# UNE INTRODUCTION AUX PROBABILITÉS

**Licence L2**
**Emmanuel Hebey**
**Année 2021-2022**

# TABLE DES MATIÈRES


**1. INTRODUCTION** **p. 04**

**PARTIE I. DÉNOMBREMENT** **p. 05**
**2. L'EXEMPLE DE GALILÉE** **p. 05**
**3. CARDINEAUX DES ENSEMBLES FINIS** **p. 05**
**4. DÉNOMBREMENT** **p. 09**
**5. QUELQUES EXERCICES 1** **p. 15**
**6. COEFFICIENTS BINOMIAUX, TRIANGLE DE PASCAL** **p. 16**
**7. FORMULE DU BINÔME** **p. 18**
**8. UNE APPLICATION AMUSANTE** **p. 20**

**PARTIE II. PROBABILITÉS DISCRÈTES FINIES** **p. 21**
**9. UNE INTRODUCTION AUX PROBABILITÉS** **p. 21**
**10. ÉLÉMENTS DE BASE DE LA THÉORIE DES ENSEMBLES** **p. 22**
**11. MATHÉMATISATION DE LA NOTION DE PROBABILITÉ** **p. 25**
**12. PROPRIÉTÉS 1** **p. 26**
**13. ÉQUIPROBABILITÉ** **p. 28**
**14. QUELQUES EXERCICES 2** **p. 29**
**15. PROPRIÉTÉS 2** **p. 32**
**16. PROBABILITÉS CONDITIONNELLES** **p. 34**
**17. FORMULE DE BAYES** **p. 36**
**18. QUELQUES EXERCICES 3** **p. 37**
**19. FORMULE DES PROBABILITÉS COMPOSÉES** **p. 38**
**20. INDÉPENDANCE** **p. 40**
**21. QUELQUES EXERCICES 4** **p. 42**

**PARTIE III. VARIABLES ALÉATOIRES** **p. 45**
**22. PREMIÈRES DÉFINITIONS** **p. 45**
**23. QUELQUES EXERCICES 5** **p. 48**
**24. DEUX LOIS DISCRÈTES FINIES USUELLES** **p. 49**

**25. LA REINE DES LOIS: LA LOI BINOMIALE** **p. 50**
**26. LA LOI HYPERGÉOMÉTRIQUE** **p. 52**
**27. MODÉLISATION ET LOIS USUELLES** **p. 52**
**28. ESPÉRANCE** **p. 53**
**29. VARIANCE ET ÉCART TYPE** **p. 56**
**30. QUELQUES EXERCICES 6** **p. 57**
**31. LE CAS DES LOIS CLASSIQUES** **p. 60**
**32. INÉGALITÉS DE BIENAYMÉ-TCHEBYCHEF ET MARKOV** **p. 63**
**33. COVARIANCE ET CORRÉLATION** **p. 66**
**34. FAMILLES DE VARIABLES ALÉATOIRES** **p. 67**
**35. QUELQUES EXERCICES 7** **p. 70**
**36. LOI FAIBLE DES GRANDS NOMBRES** **p. 72**

**PARTIE IV. LE CAS DISCRET INFINI** **p. 75**
**37. GÉNÉRALITÉS** **p. 75**
**38. LA LOI GÉOMÉTRIQUE ET LA LOI DE POISSON** **p. 75**

**PARTIE V. LOIS DES GRANDS NOMBRES** **p. 80**
**39. VARIABLES ALÉATOIRES À DENSITÉ** **p. 81**
**40. LE PROBLÈME DU CAS DISCRET** **p. 83**
**41. LOIS DES GRANDS NOMBRES** **p. 84**

**PARTIE VI. QUELQUES EXERCICES EN PLUS** **p. 87**

Licence L2 - 1er Semestre
19,5h CM

# UNE INTRODUCTION AUX PROBABILITÉS

EMMANUEL HEBEY

## 1. Introduction

En théorie des probabilités il s'agit de mathématiser les chances d'obtenir un résultat voulu. En latin "probabilis" signifie "probable", "vraisemblable", tandis que "probare" renvoie à "examiner", "vérifier". Longtemps le mot "probabilité" fut très proche du mot "chance". Antoine Augustin Cournot (1801-1877) associe encore les deux mots "probabilités et chance" dans son essai de 1843: *Exposition de la théorie des chances et des probabilités*.

La théorie des probabilités est un sujet ancien. Les premières publications sur les jeux de hasard remontent à Jérôme Cardan (1501-1576) avec son ouvrage *Liber de ludo aleae* (livre du jeu de hasard) achevé en 1564 (mais non publié avant 1663), ou encore à Johannes Kepler (1571-1630) et Galileo Galilée (1564-1642). Le véritable début de la théorie moderne remonte néanmoins aux travaux de Pierre de Fermat (1601-1665) et Blaise Pascal (1623-1662). L'étude du hasard pouvait enfin mener à quelque chose de sérieux.

Dès Galilée, dans un texte de 1612 intitulé *Considérations sur le jeux de dés*, on intuitionne la formule "magique" apprise au collège, puis au lycée,

$$\text{Probabilité} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} .$$

Il s'agira donc très vite d'ensembles et de comptage. Bien sur ce n'était pas écrit comme cela dans ce texte de Galilée. Mais il était question de la chance d'obtenir certains nombres dans le lancer de plusieurs dés, Galilée remarquant par exemple qu'il était plus avantageux dans le lancer de trois dés d'obtenir 10 par addition que 9 (il y a bien 6 cas favorables dans chacun des cas si l'on ne tient pas compte de l'ordre d'apparition des dés, mais il y a 27 cas favorables pour obtenir 10 contre 25 pour obtenir 9 si l'on tient compte de l'ordre d'apparition des dés, voir la section suivante). C'est de cela dont il sera question dans ce cours d'introduction où nous traiterons essentiellement de probabilités finies (la théorie des probabilités discrètes finies). L'objectif est d'acquérir les bases de l'étude des probabilités.

*Date*: 25 Novembre 2023.

# PARTIE I
## DÉNOMBREMENT

### 2. L'exemple de Galilée

On lance trois dés et on fait la somme des chiffres obtenus. On cherche à répondre à la question suivante: y a-t-il plus de chances d'obtenir 9 ou alors d'obtenir 10. Pour le voir on va tout simplement lister les cas de figures où l'on obtient 9, puis ensuite ceux où l'on obtient 10. Les joueurs avaient intuité qu'il était plus avantageux, dans un lancer de trois dés, de jouer le 10 plutôt que le 9. Cette interrogation était connue sous le nom de *paradoxe du grand Duc de Toscane*. Galilée répond à la question en listant effectivement les cas favorables dans les deux situations. Il obtient les deux tableaux suivants. En gras les combinaisons qui restent lorsque l'on ne distingue plus l'ordre des chiffres obtenus. Lorsque l'on veut que la somme vaille 9:

**La somme des 3 dés vaut 9**

| | | | | | |
|---|---|---|---|---|---|
| | 2-1-6 | 3-1-5 | 4-1-4 | 5-1-3 | 6-1-2 |
| **1-2-6** | **2-2-5** | 3-2-4 | 4-2-3 | 5-2-2 | 6-2-1 |
| **1-3-5** | **2-3-4** | **3-3-3** | 4-3-2 | 5-3-1 | |
| **1-4-4** | 2-4-3 | 3-4-2 | 4-4-1 | | |
| 1-5-3 | 2-5-2 | 3-5-1 | | | |
| 1-6-2 | 2-6-1 | | | | |

et lorsque l'on veut que la somme vaille 10:

**La somme des 3 dés vaut 10**

| | | | | | |
|---|---|---|---|---|---|
| | | 3-1-6 | 4-1-5 | 5-1-4 | 6-1-3 |
| | **2-2-6** | 3-2-5 | 4-2-4 | 5-2-3 | 6-2-2 |
| **1-3-6** | **2-3-5** | **3-3-4** | 4-3-3 | 5-3-2 | 6-3-1 |
| **1-4-5** | **2-4-4** | 3-4-3 | 4-4-2 | 5-4-1 | |
| 1-5-4 | 2-5-3 | 3-5-2 | 4-5-1 | | |
| 1-6-3 | 2-6-2 | 3-6-1 | | | |

Lorsque l'on ne distingue pas l'ordre d'apparition des chiffres, il y a 6 possibilités d'obtenir les sommes voulues dans chaque cas. Mais il y a par contre 27 possibilités d'obtenir 10 lorsque l'on distingue l'ordre d'apparition contre 25 possibilités d'obtenir 9. Les observations des joueurs étaient exactes: il est plus avantageux de jouer 10 plutôt que 9.

### 3. Cardinaux des ensembles finis

Un ensemble fini $E$ est un ensemble qui comporte un nombre fini d'éléments. Le cardinal d'un ensemble fini $E$ est alors le nombre d'éléments que cet ensemble comporte. On le note le plus souvent $\mathrm{Card}(E)$ ou $\sharp E$. On peut démontrer des petits résultats amusants sur cette notion. On commence par traiter du cas des ensembles en bijection. On rappelle que si $E$ et $F$ sont deux ensembles, une application $f : E \to F$ est une bijection de $E$ sur $F$ si et seulement si tout élément de $F$

possède un et exactement un antécédant par $f$. Une autre façon de dire les choses est que $f$ est bijective si et seulement si les points (i) et (ii) suivants sont vérifiés:

(i) L'application est injective, à savoir: tout élément de $F$ a au plus un antécédant par $f$, ce qui revient encore à dire que $\forall x, x' \in E$, $f(x) = f(x')$ si et seulement si $x = x'$,

(ii) L'application est surjective, à savoir: tout élément de $F$ possède au moins un antécédant par $f$, ce qui revient à dire que $\forall y \in F$, $\exists x \in E$ tel que $f(x) = y$.

Donc "$f$ bijective $\Leftrightarrow$ $f$ injective et surjective". Il existe alors une application réciproque $f^{-1} : F \to E$ telle que $f^{-1} \circ f(x) = x$ pour tout $x \in E$ et telle que $f \circ f^{-1}(y) = y$ pour tout $y \in F$. On dit que deux ensembles $E$ et $F$ sont en bijection lorsqu'il existe une bijection $f : E \to F$. La proposition suivante illustre le fait que deux ensembles en bijection ont forcément le même nombre d'éléments.

**Proposition 3.1** (Cardinal et bijections)**.** *Soient $E, F$ deux ensembles. On suppose que $E$ et $F$ sont en bijection. Alors $E$ est de cardinal fini si et seulement si $F$ est de cardinal fini et forcément $\sharp E = \sharp F$.*

*Démonstration 1:* [Longue] Notons $f : E \to F$ une bijection de $E$ sur $F$ dont l'existence est garantie par l'hypothèse que $E$ et $F$ sont en bijection. Pour démontrer ce que l'on doit démontrer on va montrer les trois points suivants:

(a) $F$ ne peut pas être infini,

(b) $F$ ne peut pas avoir moins d'éléments que $E$,

(c) $F$ ne peut pas avoir plus d'éléments que $E$.

On aura alors montré ce que l'on voulait obtenir (par symétrie du problème en $E$ et $F$ on pourrait se restreindre à ne montrer que (b) ou (c), mais démontrer (b) et (c) constitue un bon exercice de manipulation des bijections). Notons $E = \{e_1, \dots, e_n\}$ et $f(E) = \{f(e_1), \dots, f(e_n)\}$. On a clairement que $f(E) \subset F$. On démontre (a) par l'absurde. Supposons donc par l'absurde que $F$ est infini. Alors forcément il existe $y \in F \backslash f(E)$. Comme $f : E \to F$ est bijective, donc surjective, il existe $x \in E$ tel que $f(x) = y$. Il existe donc $i_0 = 1, \dots, n$ tel que $f(e_{i_0}) = y$. Ce qui est impossible puisque $y$ est par choix différent de tous les $f(e_i)$. Une contradiction et (a) est démontré. Pour obtenir (b), on note que puisque $f$ est bijective, donc injective, on a $f(e_i) \neq f(e_j)$ dès que $i \neq j$. Donc $\sharp f(E) = n$. Et comme $f(E) \subset F$, c'est que forcément $\sharp F \geq n$. En particulier $F$ ne peut pas avoir moins d'éléments que $E$, et (b) est démontré. Pour démontrer (c) on procède là encore par contradiction et exactement comme on l'a fait pour démontrer (a). On suppose par contradiction que $F$ a plus d'éléments que $E$, et donc que $\sharp F > n$. Alors forcément il existe $y \in F \backslash f(E)$ puisque $\sharp f(E) = n$ et $f(E) \subset F$. Comme $f : E \to F$ est bijective, donc surjective, il existe $x \in E$ tel que $f(x) = y$. Il existe donc $i_0 = 1, \dots, n$ tel que $f(e_{i_0}) = y$. Ce qui est impossible puisque $y$ est par choix différent de tous les $f(e_i)$. Une contradiction et (a) est démontré. Au total (a), (b), (c) sont démontrés, et comme nous l'avons déjà signalé, cela suffit à démontrer la proposition. $\square$

*Démonstration 2:* [Courte] Notons $f : E \to F$ une bijection de $E$ sur $F$ dont l'existence est garantie par l'hypothèse que $E$ et $F$ sont en bijection. Par symétrie il suffit de montrer que si $E$ est fini, alors $F$ l'est aussi et $\sharp E = \sharp F$. Comme $E$ est fini on peut écrire $E = \{e_1, \dots, e_n\}$. On a alors $f(E) = \{f(e_1), \dots, f(e_n)\}$. Comme $f$ est surjective (puisque bijective), $f(E) = F$ et $F$ et donc fini. Comme $f$

est injective (puisque bijective), les $f(e_i)$ sont deux à deux distincts. Donc $\sharp F = n$. C'est ce qu'il fallait démontrer. $\square$

On aborde maintenant le cas des produits cartésiens d'ensembles. On commence par le cas du produit de deux ensembles. Si $E$ et $F$ sont deux ensembles on rappelle que le produit cartésien de $E$ et $F$ est l'ensemble

$$E \times F = \{(e, f) \ / \ e \in E, f \in F\} \ .$$

C'est donc l'ensemble de tous les couples que l'on peut former sous la forme d'un élément de $E$ suivi d'un élément de $F$.

**Proposition 3.2** (Cardinal d'un produit)**.** *Soient $E, F$ deux ensembles finis. Alors $E \times F$ est encore un ensemble fini et $\sharp(E \times F) = \sharp E \times \sharp F$.*

*Démonstration:* Les éléments sont "comptés avec ordre", même si les deux ensembles étaient égaux. Il y a d'abord les éléments de $E$ et, ensuite, les éléments de $F$. On peut penser à un cadenas à deux chiffres tournants. Il y a d'abord le premier chiffre, puis le second (12 n'est pas 21). Comment "compter" tous les éléments de $E \times F$ ? On ordonne les éléments de $E$ et de $F$ de façon arbitraire pour pouvoir parler d'un premier élément $e_1$ de $E$, d'un second élément $e_2$ de $E$, d'un troisième élément $e_3$ de $E$ etc. (et de même pour $F$). Du coup $E = \{e_1, \ldots, e_p\}$ pour un certain $p$, qui n'est rien d'autre que le cardinal de $E$: $p = \sharp E$. De même, $F = \{f_1, \ldots, f_q\}$ et $q = \sharp F$. On a alors

$$E \times F = \{(e_i, f_j) \ / \ i = 1, \ldots, p, j = 1, \ldots, q\} \ .$$

Quand on fixe $i = 1$ il y a $q$ couples $(e_1, f_j)$. Quand on fixe $i = 2$, il y a $q$ couples $(e_2, f_j)$ etc. Il y aura donc au total $p \times q$ couples $(e_i, f_j)$ possibles ($q + q + \cdots + q$, $p$ "fois"). A savoir: $E \times F$ est un ensemble fini et $\sharp(E \times F) = \sharp E \times \sharp F$. La proposition est démontrée. $\square$

La proposition s'étend aux produits cartésiens de plus de deux ensembles. Si $E_1, \ldots, E_n$ sont $n$ ensembles, le produit cartésien $E_1 \times \cdots \times E_n$ est défini par

$$E_1 \times \cdots \times E_n = \{(e_1, \ldots, e_n) \ / \ e_1 \in E_1, \ldots, e_n \in E_n\} \ .$$

C'est donc l'ensemble de tous les $n$-uplets que l'on peut former sous la forme d'un élément de $E_1$, suivi d'un élément de $E_2$, ..., suivi d'un élément de $E_n$.

**Proposition 3.3** (Cardinal d'un produit - Suite)**.** *Soient $E_1, \ldots, E_n$ $n$ ensembles finis. Alors $E_1 \times \cdots \times E_n$ est encore un ensemble fini et $\sharp(E_1 \times \cdots \times E_n) = \sharp E_1 \times \cdots \times \sharp E_n$.*

*Démonstration.* On procède par récurrence. Une preuve par récurrence possède une *amorce* et une *hérédité*. L'amorce consiste à démontrer que la propriété voulue est vraie au premier cran (ici $n = 2$). L'hérédité consiste à démontrer que si la propriété voulue est vraie à l'ordre $n$, $n$ quelconque donné, alors elle l'est aussi à l'ordre $n + 1$. La propriété $\mathcal{P}(n)$ ici étudiée, pour $n \geq 2$, est

$$\begin{aligned} \mathcal{P}(n) \ : \ & \forall E_1, \ldots, E_n \text{ ensembles finis, } E_1 \times \cdots \times E_n \text{ est fini et} \\ & \sharp(E_1 \times \cdots \times E_n) = \sharp E_1 \times \cdots \times \sharp E_n \ . \end{aligned}$$

D'après la Proposition 3.2, $\mathcal{P}(2)$ est vraie. L'amorce est donc bien vérifiée. Supposons maintenant que pour un $n \geq 2$ fixé quelconque, la propriété $\mathcal{P}(n)$ est vraie.

Soient alors $E_1, \dots, E_{n+1}$ $n+1$ ensembles finis. Posons $E = E_1 \times \cdots \times E_n$. Il est facile de construire une bijection

$$f : E \times E_{n+1} \to E_1 \times \cdots \times E_{n+1} \ .$$

Les éléments de $E$ sont de la forme $((e_1, \dots, e_n), e_{n+1})$ où les $e_i \in E_i$. On définit alors $f$ par

$$f\left((e_1, \dots, e_n), e_{n+1}\right) = (e_1, \dots, e_{n+1})$$

pour tous $e_1 \in E_1$, $e_2 \in E_2$, ..., $e_{n+1} \in E_{n+1}$. On vérifie que $f$ est

(i) surjective: $\forall (e_1, \dots, e_{n+1}) \in E_1 \times \cdots \times E_{n+1}$, $\exists\, ((e_1, \dots, e_n), e_{n+1}) \in E \times E_{n+1}$ tel que $f\left((e_1, \dots, e_n), e_{n+1}\right) = (e_1, \dots, e_{n+1})$ .

(ii) injective: $\forall\, ((e_1, \dots, e_n), e_{n+1})\,, \left((e'_1, \dots, e'_n), e'_{n+1}\right) \in E \times E_{n+1}$, si

$$f\left((e_1, \dots, e_n), e_{n+1}\right) = f\left((e'_1, \dots, e'_n), e'_{n+1}\right) \ ,$$

alors $((e_1, \dots, e_n), e_{n+1}) = \left((e'_1, \dots, e'_n), e'_{n+1}\right)$.

Donc $f$ est bijective, et comme annoncé $E \times E_{n+1}$ et $E_1 \times \cdots \times E_{n+1}$ sont en bijection. Avec la Proposition 3.2, comme $E$ et $E_{n+1}$ sont finis, $E \times E_{n+1}$ est fini et

$$\sharp(E \times E_{n+1}) = \sharp E \times \sharp E_{n+1} \ . \tag{3.1}$$

Par hypothèse de récurrence,

$$\sharp E = \sharp E_1 \times \cdots \times \sharp E_n \ . \tag{3.2}$$

Avec la Proposition 3.1, comme $E \times E_{n+1}$ et $E_1 \times \cdots \times E_{n+1}$ sont en bijection, et comme $E \times E_{n+1}$ est fini, $E_1 \times \cdots \times E_{n+1}$ est lui aussi fini et

$$\sharp(E_1 \times \cdots \times E_{n+1}) = \sharp(E \times E_{n+1}) \ . \tag{3.3}$$

En combinant (3.1)-(3.3), on obtient que $E_1 \times \cdots \times E_{n+1}$ est fini et que

$$\sharp(E_1 \times \cdots \times E_{n+1}) = \sharp E_1 \times \cdots \times \sharp E_{n+1} \ .$$

Comme $E_1, \dots, E_{n+1}$ sont $n+1$ ensembles finis quelconques, cela signifie que la propriété $\mathcal{P}(n+1)$ est vraie elle aussi. L'hérédité est vérifiée. Cela conclue la récurrence. La proposition est démontrée. □

On termine cette section avec le cardinal d'une union de deux ensembles. Si $A$ et $B$ sont deux sous ensembles d'un ensemble $E$ (à savoir $A \subset E$ et $B \subset E$), l'union $A \cup B$ de $A$ et $B$ est le sous ensemble de $E$ constitué des $x \in E$ qui sont dans $A$ et/ou dans $B$. L'intersection $A \cap B$ de $A$ et $B$ est le sous ensemble de $E$ constitué des $x \in E$ qui sont $A$ et dans $B$ (à la fois dans $A$ et dans $B$).

**Proposition 3.4** (Cardinal d'une union)**.** *Soient $A, B$ deux sous ensembles finis d'un ensemble $E$. Alors $A \cup B$ est encore un ensemble fini et*

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

*Démonstration.* Le fait que $A \cup B$ soit fini est assez clair. Dans $\sharp A + \sharp B$ on compte deux fois les éléments de $A \cap B$. Si on retire $\sharp(A \cap B)$ à $\sharp A + \sharp B$ on va bien retrouver le nombre d'éléments de $A \cup B$. Si on veut une démonstration un peu plus rigoureuse on peut se baser sur l'observation que si $A_1, \dots, A_n$ sont des ensembles deux à deux disjoints, à savoir tels que $A_i \cap A_j = \emptyset$ pour tous $i \neq j$, alors $\sharp(A_1 \cup \cdots \cup A_n) = \sharp A_1 + \cdots + \sharp A_n$. On considère les trois ensembles $A \backslash (A \cap B)$, $A \cap B$ et $B \backslash (A \cap B)$.

Ces trois ensembles sont deux à deux disjoints. L'union de ces trois ensembles est précisément $A \cup B$. Donc

$$\begin{aligned}\sharp(A \cup B) &= \sharp(A \backslash (A \cap B)) + \sharp(A \cap B) + \sharp(B \backslash (A \cap B)) \\ &= (\sharp A - \sharp(A \cap B)) + \sharp(A \cap B) + (\sharp B - \sharp(A \cap B)) \\ &= \sharp A + \sharp B - \sharp(A \cap B)\end{aligned}$$

et on retrouve la formule voulue. La proposition est démontrée. □

La formule de la Proposition 3.4 se généralise au cas de l'union de plusieurs sous ensembles mais sous une forme peu agréable où il faut soustraire en altérné (donc altérnativement en + et −) les intersections des paires, des triplets etc. Dans le cas de trois sous ensembles $A, B, C$ on aura

$$\begin{aligned}\sharp(A \cup B \cup C) = \sharp A + \sharp B + \sharp C - \sharp(A \cap B) - \sharp(A \cap C) \\ - \sharp(B \cap C) + \sharp(A \cap B \cap C)\ .\end{aligned} \tag{3.4}$$

Les lois de Morgan, faciles à vérifier (cf. Section 10), donnent que

$$\begin{aligned}(A \cup B) \cap C &= (A \cap C) \cup (B \cap C) \\ (A \cap B) \cup C &= (A \cup C) \cap (B \cup C)\end{aligned} \tag{3.5}$$

pour tous sous ensembles $A, B, C$ d'un ensemble $E$. Pour démontrer (3.4) on écrira avec la Proposition 3.4 que

$$\begin{aligned}\sharp(A \cup B \cup C) &= \sharp(A \cup B) + \sharp C - \sharp((A \cup B) \cap C) \\ &= \sharp A + \sharp B - \sharp(A \cap B) + \sharp C - \sharp((A \cap C) \cup (B \cap C))\end{aligned} \tag{3.6}$$

puisque $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ d'après (3.5). Une nouvelle application de la Proposition 3.4 donne que

$$\sharp((A \cap C) \cup (B \cap C)) = \sharp(A \cap C) + \sharp(B \cap C) - \sharp(A \cap B \cap C) \tag{3.7}$$

puisque, comme on le constate facilement, $(A \cap C) \cap (B \cap C) = A \cap B \cap C$. En combinant (3.6) et (3.7) on obtient la formule annoncée (3.4).

La formule générale pour $n$ sous ensembles $A_1, \ldots, A_n$ est donnée par le résultat suivant.

**Lemme 3.1.** *Soient $A_1, \ldots, A_n$ des sous ensembles d'un ensemble $E$. Alors*

$$\sharp\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \sharp\left(\bigcap_{j=1}^{k} A_{i_j}\right)\ , \tag{3.8}$$

*où la seconde somme signifie que l'on somme sur tous les $k$-uplets $i_1 < i_2 < \cdots < i_k$ possibles dans $\{1, \ldots, n\}$.*

## 4. Dénombrement

Dans le langage du dénombrement il y a les permutations, les arrangements et les combinaisons. On commence avec la notion de permutation.

**Définition 4.1** (Définition d'une permutation)**.** *Soit $E$ un ensemble fini. Une permutation de $E$ est une façon de ranger les éléments de $E$. Formellement c'est une bijection de $E$ sur lui-même.*

Supposons que $E = \{1, 2\}$. Il y a deux permutations de $E$ possibles: l'identité et celle qui inverse 1 et 2. Dans le premier cas $[12] \to [12]$ et $[12] \to [21]$, cette notation abusive signifiant que $1 \to 1$ et $2 \to 2$ dans le premier cas (donc l'identité), tandis que $1 \to 2$ et $2 \to 1$ dans le second cas (inversion de 1 et de 2). Si $E = \{1, 2, 3\}$ il y a maintenant 6 permutations de $E$. On aura

$$[123] \to [123], [123] \to [213], [123] \to [132],$$
$$[123] \to [321], [123] \to [312], [123] \to [231] \ .$$

La première bijection est l'identité. Les trois suivantes sont des inversions (on inverse deux éléments et on laisse le troisième inchangé), les deux dernières sont des permutations circulaires (on fait tourner circulairement 123 pour obtenir 312, puis encore une fois pour obtenir 231). On remarque $2! = 2$ et $3! = 6$.

**Théorème 4.1** (Nombre de permutations)**.** *Il y a $n!$ permutations possibles d'un ensemble de cardinal $n$.*

*Démonstration.* On part d'un ensemble à $n$ éléments. Il faut réordonner ces $n$ éléments dans $n$ "cases" en les utilisant tous et une seule fois. Pour la première case nous avons $n$ choix possibles. Une fois l'élément de cette case choisi, il faut choisir celui de la seconde case. Là il n'y a plus que $n-1$ choix possibles puisque nous ne pouvons pas ré-utiliser l'élément de la première case. Une fois l'élément des deux premières cases choisis, il faut choisir celui de la troisième case. Là il n'y a plus que $n-2$ choix possibles puisque nous ne pouvons pas ré-utiliser l'élément de la première case, ni celui différent de la seconde case. On continue ainsi de suite pour "remplir" les $n$ cases. Le nombre de choix possibles est $n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 = n!$. Le théorème est démontré. □

**EXEMPLES D'APPLICATION:** Un anagramme est une permutation des lettres d'un mot. Il y a ainsi 120=5! anagrammes possibles du mot MATHS, et 720=6! anagrammes possibles du mot PROBAS. Une particularité de ces deux mots et qu'aucune lettre n'est répétée plusieurs fois. Si une ou plusieurs lettres sont répétées plusieurs fois le nombre d'anagrammes possibles sera le nombre d'anagrammes théorique du mot (en considérant que les lettres ne sont pas répétées) divisé par le nombre de permutations possibles des lettres répétées. Par exemple le nombre d'anagrammes possibles du mot PROBABLE où le $B$ est répété deux fois sera $8!/2! = 20.160$. Le nombre d'anagrammes possibles de ENSEMBLES où le $E$ est répété trois fois et le $S$ deux fois sera $9!/2!3! = 30.240$.

**LA NOTION DE DÉRANGEMENT:** Soit $\sigma$ une permutation de $\{1, \ldots, n\}$. On dit que $\sigma$ a un point fixe s'il existe un $i$ (ou plusieurs) tels que $\sigma(i) = i$. Si $\sigma$ n'a aucun point fixe, on parle de dérangement. Soit $D_n$ l'ensemble des dérangements de $\{1, \ldots, n\}$. Si $E = \{1, 2, 3\}$, parmi les 6 permutations de $E$ possibles, à savoir $[123] \to [123]$, $[123] \to [213]$, $[123] \to [132]$, $[123] \to [321]$, $[123] \to [312]$ et $[123] \to [231]$, seules $[123] \to [312]$ et $[123] \to [231]$ sont des dérangements. On note $E_k$ le sous ensemble des permutations de $\{1, \ldots, n\}$ qui laissent fixe $k$. Il est facile de voir que $\sharp E_k = (n-1)!$ pour tout $k$ puisque $k$ étant fixé on peut permuter les $n-1$ autres éléments comme on veut. De même, $\sharp(E_i \cap E_j) = (n-2)!$ pour tous $i \neq j$. Et, de suite en suite,

$$\sharp(E_{i_1} \cap \cdots \cap E_{i_k}) = (n-k)! \ .$$

On revient à la formule (3.8). Alors

$$\begin{aligned}\sharp(E_1 \cup \cdots \cup E_n) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \sharp \left( \bigcap_{j=1}^{k} E_{i_j} \right) \\ &= \sum_{k=1}^{n} (-1)^{k+1} (n-k)! \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} 1\end{aligned}$$

et il reste donc à compter combien il y a de $k$-uplets $i_1 < i_2 < \cdots < i_k$ possibles dans $\{1, \ldots, n\}$. Cela revient par exemple à compter les applications strictement croissantes de $\{1, \ldots, k\}$ dans $\{1, \ldots, n\}$ et donc, cf. ci-dessous, à compter les combinaisons à $k$ éléments parmi $n$. Il y a en a $\binom{n}{k}$. Donc

$$\begin{aligned}\sharp(E_1 \cup \cdots \cup E_n) &= \sum_{k=1}^{n} (-1)^{k+1} (n-k)! \binom{n}{k} \\ &= \sum_{k=1}^{n} (-1)^{k+1} \frac{n!}{k!}\end{aligned}$$

et comme $D_n = (E_1 \cup \cdots \cup E_n)^c$, on obtient que

$$\begin{aligned}\sharp D_n &= 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{n!}{k!} \\ &= n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \ ,\end{aligned}$$

une formule qui donne donc le nombre de dérangements de $\{1, \ldots, n\}$.

On aborde maintenant les notions d'arrangements. Il y en a deux: avec remise ou sans. Par famille ordonnée de $r$ éléments d'un ensemble $E$ on entendra toute "suite" de $r$ éléments de $E$ où l'ordre importe. Si par exemple $E = \{A, B, C, D, E\}$ alors $ABC$, $CBA$, $AEDC$ sont des familles ordonnées à 3 et 4 éléments de $E$, et l'on distingue donc ici $ABC$ et $CBA$ qui comportent les mêmes lettres mais dans un ordre d'apparition distinct.

**Définition 4.2** (Définition d'un arrangement)**.** *Soit $E$ un ensemble fini de $n$ éléments. Soit $r \in \mathbb{N}^\star$ un entier non nul. Un arrangement avec remise de $r$ éléments de $E$ est une famille ordonnée de $r$ éléments de $E$ où l'on s'autorise à prendre plusieurs fois un même élément de $E$. Un arrangement sans remise de $r$ éléments de $E$ est une famille ordonnée de $r$ éléments de $E$ où les éléments de $E$ sont pris une seule fois. Dans ce cas il faut exiger que $r \le n$. Un arrangement avec remise de $r$ éléments de $E$ peut-être considéré comme une application de $\{1, \ldots, r\}$ dans $E$. Un arrangement sans remise de $r$ éléments de $E$ peut-être considéré comme une injection de $\{1, \ldots, r\}$ dans $E$.*

Ici on fait le lien entre applications $f : \{1, \ldots, r\} \to E$ et familles ordonnées de $r$ éléments d'un ensemble $E$ par: si $f$ est donnée, la famille ordonnée correspondante est $f(1) \ldots f(r)$, et si c'est la famille ordonnée $e_1 \ldots e_r$ qui est donnée, alors l'application correspondante est $f(1) = e_1, \ldots, f(r) = e_r$. Les arrangements avec remise correspondent à l'expérience qui consiste à tirer un éléments d'un sac contenant $n$ éléments, puis à le remettre dans le sac, puis retirer un élément, puis le remettre dans le sac etc. $r$ fois. Les arrangements sans remise correspondent

à l'expérience qui consiste à tirer un éléments d'un sac contenant $n$ éléments, le laisser en dehors du sac, puis tirer un second élément, le laisser lui aussi en dehors du sac, etc. $r$ fois.

Considérons l'ensemble à $n = 3$ éléments $E = \{A, B, C\}$. Prenons $r = 2$ et cherchons les arrangements avec remise de 2 éléments de $E$ et les arrangements sans remise de 2 éléments de $E$. Pour les arrangements avec remise nous aurons $AA$, $AB$, $AC$, $BA$, $BB$, $BC$, $CA$, $CB$, $CC$. Il y en a $9 = 3^2$. Pour les arrangements sans remise nous aurons $AB$, $AC$, $BA$, $BC$, $CA$, $CB$. Il y en a $6 = 3!/(3-2)!$. Le choix $AA$ correspond à l'application $f : \{1, 2\} \to E$ définie par $f(1) = A$ et $f(2) = A$. Le choix $AB$ correspond à l'application $f : \{1, 2\} \to E$ définie par $f(1) = A$ et $f(2) = B$. Le choix $BA$ correspond à l'application $f : \{1, 2\} \to E$ définie par $f(1) = B$ et $f(2) = A$. La première n'est pas une injection. Les deux suivantes le sont.

**Théorème 4.2** (Nombre d'arrangements)**.** *Soit $E$ un ensemble fini de $n$ éléments. Soit $r \in \mathbb{N}^\star$ un entier non nul. Il y a $n^r$ arrangements avec remise possibles de $r$ éléments de $E$. Supposons $r \leq n$. Il y a*

$$A_n^r = \frac{n!}{(n-r)!}$$

*arrangements sans remise possibles de $r$ éléments de $E$. En particulier, il y a $n^r$ applications possibles d'un ensemble à $r$ éléments dans un ensemble à $n$ éléments, et lorsque $r \leq n$, il y a $A_n^r$ injections possibles d'un ensemble à $r$ éléments dans un ensemble à $n$ éléments (si $r > n$ aucune injection n'est possible).*

*Démonstration.* Le comptage des arrangements avec remise est simple. Il y a $r$ places à occuper et pour chaque place $n$ choix possibles. Il y a donc $n \times n \times \cdots \times n$ ($r$ fois) choix possibles, soit $n^r$. Pour les injections on raisonne de la même façon mais maintenant il est interdit d'utiliser deux fois le même élément de $E$ (ce qui est en fait la définition d'une injection). Il y a toujours $r$ places à occuper. Pour la première place il y a $n$ choix possibles. Une fois le choix pour la première place effectué, on passe à la seconde. Il y a maintenant $n - 1$ choix possibles puisqu'on ne peut pas ré-utiliser l'élément qui a été choisi pour la première place. Une fois les choix pour la première et la seconde place effectués, on passe à la troisième. Il y a maintenant $n - 2$ choix possibles puisqu'on ne peut pas ré-utiliser les éléments qui ont été choisis pour la première et la seconde place. On continue ainsi jusqu'à épuiser les $r$ places, $r \leq n$. On trouve alors $n(n-1)(n-2)\ldots(n-r+1)$ choix possibles, à savoir exactement $A_n^r$ choix possibles. Si $r > n$ il est bien sur impossible de fabriquer une famille de $r$ éléments de $E$ qui soient distincts, tout comme il est impossible qu'une application $f : \{1, \ldots, r\} \to E$ soit une injection. Il n'y a pas d'arrangements sans remise lorsque $r > n$. □

**Remarque 4.1.** *Par convention $0! = 1$. Lorsque $r = n$, $A_n^n = n!$. Les arrangements sans remise de $n$ éléments d'un ensemble à $n$ éléments correspondent effectivement aux permutations de cet ensemble.*

**EXEMPLES D'APPLICATION:** Dans un cadenas à code, contenant trois bagues tournantes de 10 chiffres (de 0 à 9), le nombre de combinaisons possibles est précisément le nombre d'arrangements avec remise à 3 éléments de l'ensemble $\{0, \ldots, 9\}$. Il y a donc $10^3 = 1.000$ combinaisons possibles. Avec 4 bagues il y

a 10.000 combinaisons possibles. Supposons maintenant que l'on veuille que le code ne comporte que des chiffres distincts. On parle alors d'arrangements sans remise. Avec 3 bagues il y a $A_{10}^3 = 720$ combinaisons possibles de chiffres distincts, et avec 4 bagues il y a $A_{10}^4 = 5.040$ combinaisons possibles de chiffres distincts.

Nous passons maintenant à la notion de combinaison. Une famille non ordonnée de $r$ éléments d'un ensemble $E$ est un sous ensemble de $E$ constitué de $r$ éléments de $E$. A partir du moment où l'on parle d'ensembles et de sous ensembles on sous entend que les éléments sont distincts. Vis à vis de la terminologie précédente, c'est un cas sans remise. Cela correspond à l'expérience qui consisterait à tirer simultanément $r$ éléments d'un sac contenant $n$ éléments disctincts.

**Définition 4.3** (Définition d'une combinaison)**.** *Soit $E$ un ensemble fini de $n$ éléments. Soit $r \in \mathbb{N}^\star$ un entier non nul avec $r \leq n$. Une combinaison de $r$ éléments de $E$ est une famille non ordonnée de $r$ éléments distincts de $E$ ou, ce qui revient au même, un sous ensemble de $E$ composé de $r$ éléments ou encore, ce qui revient toujours au même, un choix (simultané) de $r$ éléments de $E$.*

Supposons par exemple que $E = \{A, B, C, D, E\}$. Les combinaisons à 3 éléments de $E$ sont les sous ensembles (ou familles non ordonnées): $\{A, B, C\}$, $\{A, B, D\}$, $\{A, B, E\}$, $\{A, C, D\}$, $\{A, C, E\}$, $\{A, D, E\}$, $\{B, C, D\}$, $\{B, C, E\}$, $\{B, D, E\}$ et enfin $\{C, D, E\}$.

**Théorème 4.3** (Nombre de combinaisons)**.** *Soit $E$ un ensemble fini de $n$ éléments. Soit $r \in \mathbb{N}^\star$ un entier non nul avec $r \leq n$. Il y a*

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

*combinaisons de $r$ éléments de $E$. On note parfois $C_n^r = \binom{n}{r}$.*

*Démonstration.* On remarque que $\binom{n}{r} = \frac{1}{r!}A_n^r$. C'est la clef du comptage des combinaisons. Il y a $A_n^r$ familles ordonnées sans remise de $r$ éléments de $E$. Les combinaisons ne distinguent pas l'ordre des éléments. Etant donné $r$ éléments (distincts) de $E$, il y a $r!$ permutations possibles de ces éléments. Si l'on regroupe les familles ordonnées en associant entre elles celles qui ont les mêmes éléments alors on obtient un certain nombre $N$ de paquets de $r!$ familles ordonnées. Donc $A_n^r = Nr!$ et $N$ est précisément le nombre de combinaisons de $r$ éléments de $E$. Donc $N = \frac{1}{r!}A_n^r$ et le théorème est démontré. □

**Remarque 4.2.** *Une interprétation possible en termes d'applications est que l'on compte là les applications strictement croissantes de $\{1, \ldots, r\}$ dans $\{1, \ldots, n\}$. En effet, une application strictement croissante est détérminée par la donnée de son ensemble image $f(\{1, \ldots, r\}) \subset \{1, \ldots, n\}$ puisque le fait que l'application soit strictement croissante distingue un seul sous ensemble possible, celui où les éléments sont strictement ordonnés par ordre croissant. En d'autres termes, à chaque sous ensemble à $r$ éléments de $\{1, \ldots, n\}$ correspond une et une seule application strictement croissante $f : \{1, \ldots, r\} \to \{1, \ldots, n\}$. Il y a donc précisément $\binom{n}{r}$ applications strictement croissantes de $\{1, \ldots, r\}$ dans $\{1, \ldots, n\}$. C'est une des autres interprétations possibles de $\binom{n}{r}$.*

**Remarque 4.3.** *On peut introduire la notion de combinaison avec remise de $r$ éléments d'un ensemble $E$ à $n$ éléments. Il s'agit alors des familles non ordonnées*

*à $r$ éléments de $E$ où l'on s'autorise à répéter plusieurs fois les éléments de $E$. Ici $r \geq 1$ n'est plus nécessairement strictement inférieur à $n$, et cela revient maintenant à considérer des applications croissantes (non nécessairement strictement croissantes) de $\{1, \dots, r\}$ dans $\{1, \dots, n\}$. A toute famille non ordonnée de $r$ entiers dans $\{1, \dots, n\}$ correspond effectivement une seule famille où les éléments sont ordonnés de façon croissante, ce qui définit une application croissante de $\{1, \dots, r\}$ dans $\{1, \dots, n\}$. On montre qu'il y a*

$$\Gamma_n^r = \binom{n+r-1}{r} \tag{4.1}$$

*combinaisons avec remise possibles de $r$ éléments d'un ensemble $E$ à $n$ éléments. On peut démontrer ce résultat de deux façons.* **Preuve 1:** *On écrit que $E = \{e_1, \dots, e_n\}$. Le nombre de combinaisons avec remise possibles de $r$ éléments de $E$ est $\Gamma_n^r$ que l'on cherche à déterminer. On remarque qu'il y a autant de combinaisons avec remise possibles de $r$ éléments de $E$ qui ne contiennent pas $e_1$ qu'il y a de combinaisons avec remise possibles de $r$ éléments de $\{e_2, \dots, e_n\}$. Donc $\Gamma_{n-1}^r$. Et on remarque, en enlevant un $e_1$ à une combinaison avec remise de $r$ éléments de $E$ contenant au moins un $e_1$, qu'il y a autant de combinaisons avec remise possibles de $r$ éléments de $E$ qui contiennent au moins une fois $e_1$ qu'il y a de combinaisons avec remise possibles de $r-1$ éléments de $E$. Donc*

$$\Gamma_n^r = \Gamma_{n-1}^r + \Gamma_n^{r-1} \ . \tag{4.2}$$

*On a par ailleurs que $\Gamma_1^r = 1$ pour tout $r$, tandis que $\Gamma_n^0 = 1$ pour tout $n$ (l'ensemble vide étant le seul ensemble à 0 éléments qui existe). On peut maintenant démontrer la formule (4.1) par récurrence totale sur $p = n + r$, sachant que $n \geq 1$. Si $p = 1$ la formule est vraie puisqu'on va avoir affaire à $\Gamma_1^0 = 1$ et qu'il faut montrer qu'il y a égalité avec $\binom{0}{0} = 1$. On suppose maintenant que (4.1) est vraie jusqu'à un $p \geq 2$ donné quelconque, donc pour tous les $n \geq 1$ et $r$ tels que $n + r \leq p$. On considère $n \geq 1$ et $r$ quelconques tels que $n + r = p + 1$. On veut montrer (4.1) pour ces $n$ et $r$. Si $r = 0$ on a bien que $\Gamma_n^r = \binom{n+r-1}{r}$. On peut donc supposer que $r \geq 1$. En vertue de (4.2),*

$$\Gamma_n^r = \Gamma_{n-1}^r + \Gamma_n^{r-1}$$

*et $n - 1 + r = n + r - 1 \leq p$. Par hypothèse de récurrence on a donc que*

$$\begin{aligned}
\Gamma_n^r &= \frac{(n+r-2)!}{r!(n-2)!} + \frac{(n+r-2)!}{(r-1)!(n-1)!} \\
&= \frac{(n+r-2)!}{r!(n-1)!}\,(n-1+r) \\
&= \frac{(n+r-1)!}{r!(n-1)!}
\end{aligned}$$

*et on récupère la formule (4.1). La formule est ainsi démontrée par récurrence totale. Une récurrence totale fonctionne de la façon suivante: elle possède une amorce et une hérédité. L'amorce consiste à démontrer que la propriété voulue est vraie au premier cran. L'hérédité consiste à démontrer que si la propriété voulue est vraie jusqu'à l'ordre $n$, $n$ quelconque donné, alors elle l'est aussi à l'ordre $n+1$.* **Preuve 2:** *Il y a cependant une preuve beaucoup beaucoup plus simple si l'on décide de compter les applications croissantes de $\{1, \dots, r\}$ dans $\{1, \dots, n\}$. Si $\hat{f}$ est une*

*telle application, on définit $f$ par $f(k) = \hat{f}(k) + k - 1$. On a $f(1) = \hat{f}(1)$. On remarque que $\hat{f}$ croissante équivaut à $f$ strictement croissante, car*

$$f(k+1) - f(k) = \hat{f}(k+1) - \hat{f}(k) + 1 \ .$$

*On remarque que $f$ est maintenant une application de $\{1, \ldots, r\}$ à valeurs dans $\{1, \ldots, n+r-1\}$. On sait compter les applications strictement croissantes de $\{1, \ldots, r\}$ dans $\{1, \ldots, n+r-1\}$. Il y en a $\binom{n+r-1}{r}$. D'où la relation (4.1).*

**EXEMPLES D'APPLICATION:** Un jeu de cartes contient 32 cartes. On tire 5 cartes. Il y a alors $\binom{32}{5} = 201.376$ combinaisons possibles de 5 cartes parmi ces 32 cartes. Au loto on tire 6 numéros parmi 49. Il y a alors $\binom{49}{6} = 13.983.816$ tirages possibles.

## 5. Quelques exercices 1

**EXERCICE 1:** Une course d'athlétisme comprend 35 participants. Le premier, le second et le troisième de la course sont distingués par une médaille d'or, d'argent et de bronze. Il constituent le podium de la course. Combien a priori y a-t-il de podiums possibles ?

CORRECTION: Typiquement il s'agit de savoir combien de familles ordonnées sans remise de 3 participants peut-on former. La réponse et $A_{35}^3 = 39.270$.

**EXERCICE 2:** En informatique, on utilise le système binaire pour coder les caractères. Un bit est un élément qui prend la valeur 0 ou la valeur 1. Un octet est composé de 8 bits. Combien de caractères peut-on coder avec un octet ?

CORRECTION: Typiquement il s'agit de savoir combien de familles ordonnées avec remise de 8 éléments peut-on former avec des 0 et des 1. La réponse et $2^8 = 256$.

**EXERCICE 3:** Le clavier d'une porte codée contient 10 chiffres $0, 1, \ldots, 9$ et 6 lettres $A, B, C, D, E, F$. Le code d'entrée est constituée d'une lettre suivie de 3 chiffres. Combien y a-t-il de codes possibles ? Et si le code d'entrée est constitué d'une lettre et de 3 chiffres, peu importe où apparait la lettre ?

CORRECTION: Dans le premier cas il s'agit de choisir une lettre parmi 6, il y a donc 6 choix, et de choisir 3 chiffres parmi 10, de façon ordonnée avec remise puisqu'un chiffre peut être utilisé plusieurs fois (on peut avoir comme code A001 par exemple). Il y a donc $6 \times 10^3 = 6.000$ choix possibles. Dans le second système de codes, la lettre peut être à la première place, à la seconde, à la troisième, ou à la quatrième et dernière place. Dans chacun des cas il y a 6.000 choix possibles. Il y aura donc $4 \times 6.000 = 24.000$ choix possibles.

**EXERCICE 4:** Dans une classe de 32 élèves, on compte 19 garçons et 13 filles. On doit élire deux délégués. Quel est le nombre de choix possibles ? Quel est le nombre de choix si l'on impose un garçon et une fille ? Quel est le nombre de choix si l'on impose 2 garçons ?

CORRECTION: Dans le premier cas il s'agit de dénombrer les familles non ordonnées de 2 élèves parmi 32 élèves. On trouve qu'il y en a $\binom{32}{2} = 496$. Dans le second cas il faut choisir 1 délégué parmi les garçons, il y a 19 choix possibles, et une déléguée parmi les filles, il y a 13 choix possibles. Il y aura donc $19 \times 13 = 247$

choix possibles. Dans le dernier cas il faut dénombrer les familles non ordonnées de 2 élèves parmi 19 garçons. On trouve qu'il y en a $\binom{19}{2} = 171$.

**EXERCICE 5⋆:** Soit $X$ l'ensemble des nombres à 7 chiffres qui ne comportent aucun 1, $X_1 \subset X$ le sous ensemble de $X$ des nombres qui comportent 7 chiffres différents, $X_2 \subset X$ le sous ensemble de $X$ des nombres pairs, $X_3 \subset X$ le sous ensemble de $X$ des nombres dont les chiffres forment une suite strictement croissante en lecture de gauche à droite. Donner les cardinaux de ces différents ensembles. Attention un nombre ne commence pas par un 0.

CORRECTION: En ce qui concerne $X$ il y a 8 choix pour le premier chiffre (de 2 à 9), et il y a 9 choix pour chacun des 6 autres chiffres (de 0 à 9 sauf 1). On trouve donc $\sharp X = 8 \times 9^6$. En ce qui concerne $X_1$ il y a toujours 8 choix pour le premier chiffre. Une fois ce chiffre choisi on veut dénombrer les familles ordonnées sans remise de 6 chiffres dans une ensemble de 8 chiffres (9 moins le premier chiffre choisi). Il y en a $A_8^6$. On a donc $\sharp X_1 = 8 \times A_8^6$. En ce qui concerne $X_2$ il y a 5 façons de choisir le 7ème chiffre qui doit être l'un des chiffres 0, 2, 4, 6, 8. Il y a toujours 8 choix possibles pour le premier chiffre et 9 choix pour les chiffres de la 2nde à la 6ème place. On a donc $\sharp X_2 = 8 \times 9^5 \times 5 = 40 \times 9^5$. Reste à calculer le cardinal de $X_3$. On ne trouvera jamais le chiffre 0 puisqu'il ne peut pas figurer en première place ni ensuite dans une liste strictement ordonnée par ordre croissant de gauche à droite. On cherche donc les applications strictement croissantes de $\{1, \ldots, 7\}$ dans $\{2, \ldots, 9\}$ (8 éléments). Il y en a donc $\binom{8}{7} = 8$ (cf. la Remarque 4.2 ci-dessus). Soit $\sharp X_3 = 8$.

**EXERCICE 6⋆:** On souhaite ranger sur une étagère 4 livres de mathématiques (distincts), 6 de physique (distincts), et 3 de chimie (distincts). De combien de façons peut-on effectuer ce rangement si les livres doivent être groupés par matières ? Et si seuls les livres de mathématiques doivent être groupés ensemble ?

CORRECTION: Il y a 3 matières, et donc $3! = 6$ façon d'ordonner les matières. Il y a par ailleurs $4! = 24$ façons d'ordonner les livres de mathématiques, $6! = 720$ façons d'ordonner les livres de physique et $3! = 6$ façons d'ordonner les livres de chimie. Au total il y a $6 \times (24 \times 720 \times 6) = 622.080$ façons d'effectuer le rangement si les livres doivent être groupés par matières. Si seuls les livres de mathématiques doivent être groupés ensemble, il y a $0, 1, \ldots, 9$ livres possibles avant que le groupe des livres de mathématiques n'apparaisse, soit 10 configurations possibles. Il y a toujours $4!$ façons de ranger les livres de mathématiques, et $9!$ façons de ranger les autres livres. Il y a donc $10 \times 4! \times 9! = 87.091.200$ façons d'effectuer le rangement si seuls les livres de mathématiques doivent être groupés ensemble

## 6. Coefficients binomiaux, triangle de Pascal

Les $\binom{n}{p}$ sont aussi appelés coefficients binomiaux. On rappelle que

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} \ . \tag{6.1}$$

Ils vérifient un certain nombre de propriétés. On vérifie sans aucune difficulté (et sans calcul aucun) à partir de la formule ci-dessus que:

$$\begin{aligned} \forall\ 0 \leq p \leq n, \binom{n}{p} &= \binom{n}{n-p}\ , \\ \forall\ 1 \leq p \leq n, \binom{n}{p} &= \frac{n}{p}\binom{n-1}{p-1}\ . \end{aligned} \tag{6.2}$$

La seconde des relations (6.2) est dite relation d'absorption. Une autre relation liée à la loi hypergéométrique, et appelée identité de Vandermonde, est aussi énoncée en (31.5). La plus importante des relations sur les coefficients binomiaux est sans aucun doute la formule dite de Pascal qui est à la base du triangle du même nom. On énonce cette formule de Pascal dans le lemme suivant.

**Lemme 6.1** (Formule de Pascal)**.** *On a*

$$\binom{n}{p} + \binom{n}{p+1} = \binom{n+1}{p+1}$$

*pour tous* $0 \leq p \leq n-1$.

*Démonstration.* On calcule à partir de la formule (6.1). On a

$$\begin{aligned} \binom{n}{p} + \binom{n}{p+1} &= \frac{n!}{p!(n-p)!} + \frac{n!}{(p+1)!(n-p-1)!} \\ &= \frac{n!}{(p+1)!(n-p)!}\,(p+1+n-p)) \\ &= \frac{(n+1)!}{(p+1)!(n-p)!} \end{aligned}$$

et on retrouve bien la formule de $\binom{n+1}{p+1}$. Le lemme est démontré. □

**Proposition 6.1.** *Pour tout* $n \in \mathbb{N}^\star$ *et tout* $p \in \{0, 1, \ldots, n\}$, $\binom{n}{p}$ *est un entier naturel. De plus, si* $p \geq 1$ *et si* $n$ *et* $p$ *sont premiers entre eux (donc aussi* $p < n$ *car sinon* $n$ *et* $p$ *ne sont pas permiers entre eux), alors* $n$ *divise* $\binom{n}{p}$.

*Démonstration.* On commence déjà par montrer que pour tout $n \in \mathbb{N}^\star$ et tout $p \in \{0, 1, \ldots, n\}$, $\binom{n}{p}$ est un entier naturel. On peut tout simplement invoquer le fait que $\binom{n}{p}$ compte les combinaisons, et donc est forcément entier. On peut aussi proposer une autre preuve en effectuant une récurrence forte sur $n$ à partir de la formule de Pascal. Pour $n \in \mathbb{N}^\star$, soit $(\mathcal{P}_n)$ la propriété:

$$(\mathcal{P}_n):\ \forall p \in \{0, 1, \ldots, n\}, \binom{n}{p} \in \mathbb{N}\ .$$

Clairement $(\mathcal{P}_1)$ est vraie. Soit $n \in \mathbb{N}^\star$. On suppose que $(\mathcal{P}_k)$ est vraie pour tout $0 \leq k \leq n$. On montre que $(\mathcal{P}_{n+1})$ est vraie, donc que $\binom{n+1}{p} \in \mathbb{N}$ pour tout $0 \leq p \leq n+1$. Si $p = 0$ ou $p = n+1$ on a clairement $\binom{n+1}{p} \in \mathbb{N}$. Supposons $1 \leq p \leq n$. Avec la formule de Pascal:

$$\binom{n}{p-1} + \binom{n}{p} = \binom{n+1}{p}$$

et par récurrence forte $\binom{n+1}{p}$ est un entier puisque somme de deux entiers. Donc $(\mathcal{P}_{n+1})$ est vraie, ce qui achève la récurrence. On aborde maintenant la preuve de

la seconde partie de la proposition. Il s'agit d'une conséquence directe de la formule d'absorption. On a

$$n\binom{n-1}{p-1} = p\binom{n}{p} \ .$$

Donc, comme les coefficients binomiaux sont des entiers, $n$ divise $p\binom{n}{p}$ au sens de la division euclidienne ($a$ divise $b$ ssi $b$ est un mutiple de $a$). Comme $n$ et $p$ sont premiers entre eux on peut invoquer le théorème de Gauss ou remarquer qu'aucun des éléments de la décomposition en nombre premiers de $n$ ne peut se retrouver dans $p$ de sorte qu'ils se retrouvent tous dans $\binom{n}{p}$. En d'autres termes, $n$ divise $\binom{n}{p}$ comme annoncé. La proposition est démontrée. $\square$

La seconde partie de la proposition s'améliore facilement en la propriété que pour tout $n \in \mathbb{N}^\star$ et tout $1 \leq p \leq n$, l'entier $\frac{n}{\mathrm{pgcd}(n,p)}$ divise $\binom{n}{p}$, où $\mathrm{pgcd}(n,p)$ représente le PGCD de $n$ et $p$.

Le triangle de Pascal est une façon de placer en triangle les coefficients binomiaux afin de les calculer facilement. Il est basé sur la formule du Lemme 6.1. Le numéro des lignes correspond au $n$ du $\binom{n}{p}$ sachant que l'on commence à la ligne 0 (le sommet du triangle). Et quand on progresse de gauche à droite dans la ligne $n$ on trouve $\binom{n}{0}$, $\binom{n}{1}$, $\binom{n}{2}$ etc. Les bordures du triangles de Pascal valent toutes 1 et correspondent aux $\binom{n}{0}$ et $\binom{n}{n}$ (qui valent effectivement toujours 1). A l'intérieur du triangle, chaque élément est obtenu comme la somme des deux éléments de la ligne précédente qui sont juste au dessus de lui, ce qui correspond à la formule du Lemme 6.1. Avec les calculatrices modernes cette construction en triangle permettant de calculer les coefficients du binôme (en construisant le triangle de ligne en ligne en partant du sommet) a un peu perdu d'intérêt. Mais il reste dans la culture mathématique et un passage obligé lorsque l'on parle de coefficients binomiaux.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1

LE TRIANGLE DE PASCAL

## 7. Formule du binôme

La formule du binôme de Newton dont nous allons maintenant parler est la généralisation à la puissance $n$ des formules $(a+b)^2 = a^2 + 2ab + b^2$ pour $n = 2$ et $(a+b)^3 = a^3 + 3ab^2 + 3a^2b + b^3$ pour $n = 3$. La formule est importante. Sa preuve est basée sur la formule de Pascal du Lemme 6.1.

**Théorème 7.1** (Formule du binôme)**.** *Pour tout entier $n$ et tous réels $a, b \in \mathbb{R}$,*

$$(a+b)^n = \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p} \tag{7.1}$$

*où les $\binom{n}{p}$ sont les coefficients binomiaux donnés par* (6.1)*.*

On calcule $\binom{2}{0} = 1$, $\binom{2}{1} = 2$ et $\binom{2}{2} = 1$. On retrouve bien la formule $(a+b)^2 = a^2 + 2ab + b^2$. On calcule $\binom{3}{0} = 1$, $\binom{3}{1} = 3$, $\binom{3}{2} = 3$ et $\binom{3}{3} = 1$. On retrouve bien la formule $(a+b)^3 = a^3 + 3ab^2 + 3a^2b + b^3$.

*Démonstration.* On procède par récurrence sur $n$. L'amorce pour $n = 1$ est facile à vérifier puisque $\binom{1}{0} = \binom{1}{1} = 1$. Pour s'attaquer à l'hérédité de la preuve par récurrence, on suppose que la formule (7.1) est vraie à l'ordre $n$. On veut la démontrer à l'ordre $n+1$. On écrit pour commencer que

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n \\
&= (a+b) \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p} \\
&= \sum_{p=0}^{n} \binom{n}{p} a^{p+1} b^{n-p} + \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p+1} \ .
\end{aligned}$$

On change la sommation dans la première des sommes en décalant d'un le compteur $p$ pour ramener $p+1$ à $p$:

$$\sum_{p=0}^{n} \binom{n}{p} a^{p+1} b^{n-p} = \sum_{p=1}^{n+1} \binom{n}{p-1} a^p b^{n-p+1} \ .$$

On récupère alors que

$$\begin{aligned}
(a+b)^{n+1} &= \sum_{p=1}^{n+1} \binom{n}{p-1} a^p b^{n-p+1} + \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p+1} \\
&= \sum_{p=1}^{n} \left( \binom{n}{p} + \binom{n}{p-1} \right) a^p b^{n-p+1} + a^{n+1} + b^{n+1}
\end{aligned}$$

lorsque l'on "isole" les extrémités $p = n+1$ dans la première somme et $p = 0$ dans la seconde somme pour mettre ensemble les deux sommes de 1 à $n$ restantes. En vertue de la formule de Pascal,

$$\binom{n}{p} + \binom{n}{p-1} = \binom{n+1}{p} \ .$$

Donc

$$\begin{aligned}
(a+b)^{n+1} &= \sum_{p=1}^{n} \left( \binom{n}{p} + \binom{n}{p-1} \right) a^p b^{n-p+1} + a^{n+1} + b^{n+1} \\
&= \sum_{p=1}^{n} \binom{n+1}{p} a^p b^{n-p+1} + a^{n+1} + b^{n+1} \\
&= \sum_{p=0}^{n+1} \binom{n+1}{p} a^p b^{n-p+1}
\end{aligned}$$

lorsque l'on ré-intègre les termes $a^{n+1}$ et $b^{n+1}$ à la somme, et la formule (7.1) est donc aussi vraie à l'ordre $n+1$. Cela achève la récurrence. Le théorème est démontré. $\square$

## 8. Une application amusante

Que peut-on faire avec la formule du binôme ? Plein de choses, ce qui n'est pas très surprenant puisqu'elle généralise les formules pour $(a+b)^2$ et $(a+b)^3$ déjà rencontrées maintes et maintes fois dans vos études. Une application en théorie des probabilités tient en la loi binomiale qui sera discutée plus loin dans ce polycopié. On sort brièvement du monde des probabilités dans cette courte section pour démontrer le petit théorème de Fermat, qui appartient à l'univers arithmétique, avec la formule du binôme. On rappelle la définition de la congruence: si $a$ et $b$ sont deux entiers, et $n$ en est un troisième, on dit que $a$ est congru à $b$ modulo $n$, et on écrit $a \equiv b\ (n)$, si (pour siginifier que) $a-b$ est multiple de $n$. Donc $a \equiv b\ (n)$ si et seulement si $\exists k \in \mathbb{Z}$ tel que $a-b=kn$.

**Théorème 8.1** (Petit théorème de Fermat)**.** *Pour tout entier $n \in \mathbb{N}^\star$ et tout nombre premier $p$, $n^p \equiv n\ (p)$.*

*Démonstration.* A titre de relarque, si $p=2$ le petit théorème de Fermat est simple a établir puisqu'il dit juste que pour tout $n$, $n^2-n$ est pair. C'est une évidence puisqu'un entier est soit pair soit impair et que le carré d'un pair est pair tandis que le carré d'un impair est impair (et que la différence de deux entiers ayant la même parité est toujours un nombre pair). On fixe maintenant $p$ premier, par exemple tel que $p \geq 3$. On montre par récurrence sur $n$ que pour tout $n \geq 1$, $n^p \equiv n\ (p)$. Si $n=1$ c'est une évidence puisque $1^p-1=0$. On suppose que pour un $n$, $n^p \equiv n\ (p)$. Avec la formule du binôme,

$$\begin{aligned}(n+1)^p &= \sum_{k=0}^{p} \binom{p}{k} n^k \\ &= 1+n^p+\sum_{k=1}^{p-1} \binom{p}{k} n^k\ .\end{aligned} \tag{8.1}$$

Par hypothèse de récurrence $n^p \equiv n\ (p)$. Par ailleurs, comme $p$ est premier, les entiers $k$ et $p$ sont premiers entre eux pour tout $1 \leq k \leq p-1$. En raison de la proposition 6.1 on a donc que $\binom{p}{k}$ est un multiple de $p$ pour tout $1 \leq k \leq p-1$. En particulier la somme dans la seconde ligne de (8.1) est un multiple de $p$. Ainsi

$$(n+1)^p = 1+n+ \text{ un multiple de } p$$

et donc $(n+1)^p \equiv n+1\ (p)$. La récurrence est achevée. Le petit théorème de Fermat est démontré. $\square$

# PARTIE II
## PROBABILITÉS DISCRÈTES FINIES

### 9. Une introduction aux probabilités

Une expérience aléatoire est une expérience dont on ne peut pas prévoir le résultat de façon certaine. C'est typiquement le cas du lancer d'un ou de plusieurs dés, ou lorsque l'on tire des cartes, ou lorsque l'on lance une pièce à pile ou face. Dans une expérience aléatoire la première chose que l'on fait est de trouver l'ensemble de tous les résultats possibles. C'est l'ensemble $\Omega$ sur lequel on va pouvoir définir une probabilité. Dans le lancer d'une pièce, $\Omega = \{\text{pile}, \text{face}\}$. Dans le lancer d'un dé $\Omega_1 = \{1, 2, \ldots, 6\}$. Dans le lancer de deux dés, si on note les résultats obtenus, $\Omega_2 = \{(i, j) \ / \ i = 1, \ldots, 6, j = 1, \ldots, 6\}$. Dans le lancer de trois dés pour revenir à l'étude de Galilée,

$$\Omega_3 = \{(i, j, k) \ / \ i = 1, \ldots, 6, j = 1, \ldots, 6, k = 1, \ldots, 6\} \ .$$

Dans ces deux derniers cas, il s'agit de couples ou triplets ordonnés, puisque mathématiquement $(1, 2) \neq (2, 1)$. Dans l'expérience ici envisagée on lance donc tout d'abord un dé, puis un autre, etc. ou alors les dés sont de couleurs différentes que l'on distingue. On pourrait considérer que l'ordre n'importe pas auquel cas, mathématiquement, l'écriture de $\Omega$ choisie est incorrecte et il faudrait plutôt écrire $\Omega = \{\{i, j\} \ / \ i = 1, \ldots, 6, j = 1, \ldots, 6\}$ dans le cas de deux dés et, dans le cas de trois dés, $\Omega = \{\{i, j, k\} \ / \ i = 1, \ldots, 6, j = 1, \ldots, 6, k = 1, \ldots, 6\}$.

D'un point de vue intuitif, un événement $A$ est une affirmation qui peut être vraie ou fausse selon le résultat de l'expérience. Dans le cas du lancer d'un dé des événements types pourraient être:

(i) $A_1$ le résultat obtenu est 3;

(ii) $A_2$ le résultat obtenu est pair.

Dans le cas de l'expérience de Galilée du lancer de trois dés, les deux événements considérés sont:

(iii) $A_3$ la somme des trois dés vaut 9;

(iv) $A_4$ la somme des trois dés vaut 10.

D'un point de vue ensembliste, dans les cas (i) et (ii), l'ensemble de tous les résultats possibles, l'univers des possibles, est l'ensemble $\Omega_1$ ci-dessus. Et on peut regarder $A_1$ et $A_2$ comme les sous ensembles de $\Omega_1$ donnés par $A_1 = \{3\}$ et $A_2 = \{2, 4, 6\}$. Dans les cas (iii) et (iv), l'ensemble de tous les résultats possibles, l'univers des possibles, est l'ensemble $\Omega_3$ ci-dessus. Et on peut regarder $A_3$ et $A_4$ comme les sous ensembles de $\Omega_3$ donnés par les tableaux de la Section 2, à savoir

$$\begin{aligned} A_3 = \Big\{ & (1,2,6), (1,3,5), (1,4,4), (1,5,3), (1,6,2), (2,1,6), (2,2,5), (2,3,4), (2,4,3), \\ & (2,5,2), (2,6,1), (3,1,5), (3,2,4), (3,3,3), (3,4,2), (3,5,1), (4,1,4), (4,2,3), \\ & (4,3,2), (4,4,1), (5,1,3), (5,2,2), (5,3,1), (6,1,2), (6,2,1) \Big\} \end{aligned}$$

et

$$A_4 = \Big\{(1,3,6),(1,4,5),(1,5,4),(1,6,3),(2,2,6),(2,3,5),(2,4,4),(2,5,3),(2,6,2),$$
$$(3,1,6),(3,2,5),(3,3,4),(3,4,3),(3,5,2),(3,6,1),(4,1,5),(4,2,4),(4,3,3),$$
$$(4,4,2),(4,5,1),(5,1,4),(5,2,3),(5,3,2),(5,4,1),(6,1,3),(6,2,2),(6,3,1)\Big\} .$$

La formule "magique" des probabilités discrètes finies que vous avez étudiée au collège et au lycée est

$$\text{Probabilité} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} .$$

Le nombre de cas possibles est le cardinal de l'univers des possibles $\Omega$ et le nombre de cas favorables est le cardinal de l'événement $A$. Ici $\sharp\Omega_1 = 6$, $\sharp A_1 = 1$ et $\sharp A_2 = 3$. La probabilité d'obtenir 3 dans le lancé d'un dé, on notera $P(A_1)$, vaut donc $P(A_1) = \frac{1}{6}$. De même, la probabilité d'obtenir un nombre pair dans le lancé d'un dé, on notera $P(A_2)$, vaut ainsi $P(A_2) = \frac{3}{6} = \frac{1}{2}$. On a $P(A_2) > P(A_1)$. Il y a plus de chances d'obtenir un nombre pair lors du lancer d'un dé que le chiffre 3. Le cardinal de $\Omega_3$, puisqu'il s'agit de familles ordonnées avec remises (un même chiffre peut apparaître plusieurs fois) est donné par $\sharp\Omega_3 = 6^3 = 216$ (cf. Section 4). On a par ailleurs que $\sharp A_3 = 25$ tandis que $\sharp A_4 = 27$. La probabilité d'obtenir la somme de 9 avec le lancé de 3 dés, on notera $P(A_3)$, vaut donc

$$P(A_3) = \frac{\sharp A_3}{\sharp\Omega_3} = \frac{25}{216}$$

La probabilité d'obtenir la somme de 10 avec le lancé de 3 dés, on notera $P(A_4)$, vaut quant à elle

$$P(A_4) = \frac{\sharp A_4}{\sharp\Omega_3} = \frac{27}{216} = \frac{1}{8} .$$

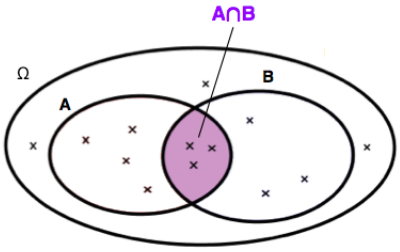
On a $P(A_4) > P(A_3)$. Il y a plus de chances d'obtenir une somme de 10 lors du lancer de 3 dés qu'une somme de 9. La formule pour $P$ utilisée ci-dessus correspond au cas d'équiprobabilité défini par le fait que les éléments de $\Omega$ (les événements élémentaires) ont tous la même probabilité (la même chance) de se produire.

## 10. Éléments de base de la théorie des ensembles

Soit $\Omega$ un ensemble, et soient $A \subset \Omega$ et $B \subset \Omega$ deux sous ensembles de $\Omega$. On définit l'intersection $A \cap B$ de $A$ et $B$ comme étant le sous ensemble de $\Omega$, possiblement l'ensemble vide $\emptyset$, constitué des éléments de $\Omega$ qui sont à la fois dans $A$ et dans $B$. Il est donné par

$$A \cap B = \{x \in \Omega \ / \ x \in A \text{ et } x \in B\} .$$

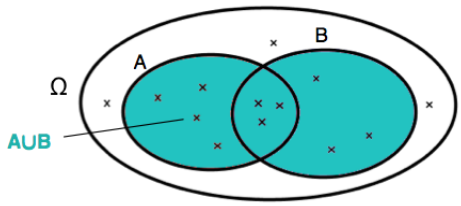
On a donc:



On définit l'union $A \cup B$ de $A$ et $B$ comme étant le sous ensemble de $\Omega$ constitué des éléments de $\Omega$ qui sont dans $A$ et/ou dans $B$. Il est donné par

$$A \cup B = \{x \in \Omega \ / \ x \in A \text{ ou } x \in B\} \ .$$

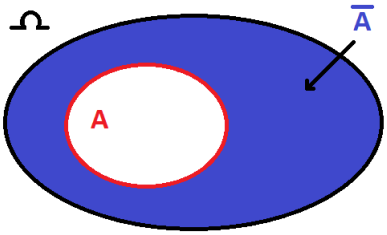
On a donc:




On définit le complémentaire de $A$ dans $\Omega$, on note souvent $A^c$, comme étant le sous ensemble de $\Omega$ constitué des éléments de $\Omega$ qui ne sont pas dans $A$. Il est donné par

$$A^c = \{x \in \Omega \ / \ x \notin A\} \ .$$

On a donc:




Les premières formules de Morgan gèrent la distributivité de l'intersection sur l'union et récirproquement.

**Lemme 10.1** (Formules de Morgan 1)**.** *Soit $\Omega$ un ensemble. On a:*

*(1)* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ ,

*(2)* $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

*pour tous $A, B, C \subset \Omega$ sous ensembles de $\Omega$.*

*Démonstration.* L'égalité de deux sous ensembles se démontre par double inclusion. Soient $A, B, C \subset \Omega$ trois sous ensembles quelconques de $\Omega$. Pour montrer (1) on devra donc montrer:

(1a) $(A \cup B) \cap C \subset (A \cap C) \cup (B \cap C)$ ,

(1b) $(A \cap C) \cup (B \cap C) \subset (A \cup B) \cap C$ .

L'inclusion $(A \cap C) \cup (B \cap C) \subset (A \cup B) \cap C$ est immédiate puisque $A \cap C \subset A$ et $B \cap C \subset B$ entraînent que $A \cap C \subset A \cup B$ et $B \cap C \subset A \cup B$, tandis que $A \cap C \subset C$ et $B \cap C \subset C$ entraînent à leur tour que $A \cap C \subset (A \cup B) \cap C$ et $B \cap C \subset (A \cup B) \cap C$. Donc (1b) est vrai. Pour montrer (1a) on considère $x \in (A \cup B) \cap C$, $x$ quelconque. Alors $x \in A \cup B$ et $x \in C$. Comme $x \in A \cup B$, forcément $x \in A$ ou alors $x \in B$ (et on peut même avoir les deux bien sur). Supposons $x \in A$. Alors $x \in A \cap C$ puisqu'on a toujours $x \in C$. De même, si $x \in B$ alors $x \in B \cap C$ puisqu'on a toujours $x \in C$. Donc, quelque soit le cas de figure, $x \in (A \cap C) \cup (B \cap C)$. Donc

tout élément $x$ de $x \in (A \cup B) \cap C$ est dans $(A \cap C) \cup (B \cap C)$. On a donc aussi démontré (1a). Donc (1) est démontré. Pour démontrer (2) il faut montrer que

(2a) $(A \cap B) \cup C \subset (A \cup C) \cap (B \cup C)$ ,

(2b) $(A \cup C) \cap (B \cup C) \subset (A \cap B) \cup C$ .

L'inclusion (2a) est immédiate puisque $A \cap B \subset A$ de sorte que $(A \cap B) \cup C \subset A \cup C$ tandis que $A \cap B \subset B$ entraîne lui que $(A \cap B) \cup C \subset B \cup C$. D'où (2a). Pour montrer (2b) on considère $x \in (A \cup C) \cap (B \cup C)$, $x$ quelconque. Si $x \in C$ alors clairement $x \in (A \cap B) \cup C$. Il faut donc maintenant traiter le cas $x \notin C$. Supposons donc $x \notin C$. On a $x \in A \cup C$ donc forcément $x \in A$. De même, $x \in B \cup C$ et donc forcément $x \in B$. Donc $x \in A \cap B$ et, en particulier, $x \in (A \cap B) \cup C$. Donc tout élément $x \in (A \cup C) \cap (B \cup C)$ est dans $(A \cap B) \cup C$. On a donc aussi démontré (2b). Donc (2) est démontré. Le lemme est démontré. □

Les secondes formules de Morgan gèrent l'action du complémentaire sur les unions et intersections. En peu de mots: le complémentaire de l'union est égal à l'intersection des complémentaires. Le complémentaire de l'intersection est égal à l'union des complémentaires.

**Lemme 10.2** (Formules de Morgan 2)**.** *Soit $\Omega$ un ensemble. On a:*

*(1)* $(A \cup B)^c = A^c \cap B^c$ ,

*(2)* $(A \cap B)^c = A^c \cup B^c$

*pour tous $A, B \subset \Omega$ sous ensembles de $\Omega$.*

*Démonstration.* Là encore il faut démontrer des doubles inclusions. Soient $A, B \subset \Omega$ sous ensembles de $\Omega$. Pour montrer (1) il faut ainsi montrer que:

(1a) $(A \cup B)^c \subset A^c \cap B^c$ ,

(1b) $A^c \cap B^c \subset (A \cup B)^c$ .

On a clairement $(A \cup B)^c \subset A^c$ et $(A \cup B)^c \subset B^c$. Donc $(A \cup B)^c \subset A^c \cap B^c$ et (1a) est démontré. Pour démontrer (1b) on considère $x \in A^c \cap B^c$, $x$ quelconque. Alors $x \notin A$ et $x \notin B$. Donc $x \notin A \cup B$, soit $x \in (A \cup B)^c$ et ainsi, puisque $x$ est quelconque, (1b) est aussi démontré. On a donc bien (1). On procède de la même manière pour démontrer (2). Il nous faut montrer que:

(2a) $(A \cap B)^c \subset A^c \cup B^c$ ,

(2b) $A^c \cup B^c \subset (A \cap B)^c$ .

Comme $A^c \subset (A \cap B)^c$ et $B^c \subset (A \cap B)^c$ on a clairement que (2b) est vrai. Pour démontrer (2a) on considère $x \in (A \cap B)^c$. Si $x \notin A$ alors $x \in A^c$ et donc $x \in A^c \cup B^c$. Si $x \in A$ alors forcement $x \notin B$ car $x \notin A \cap B$, et donc $x \in B^c$. Là encore, $x \in A^c \cup B^c$. Comme $x$ est quelconque, (2a) est démontré. On a donc aussi (2). Le lemme est démontré. □

On généralise l'intersection de deux sous ensembles à l'intersection de plusieurs sous ensembles $A_1, \dots, A_n$ par

$$A_1 \cap \cdots \cap A_n = \{x \in \Omega \ / \ x \in A_i, \forall i = 1, \dots, n\} \ ,$$

et on généralise l'union de deux sous ensembles à l'union de plusieurs sous ensembles $A_1, \dots, A_n$ par

$$A_1 \cup \cdots \cup A_n = \{x \in \Omega \ / \ \exists i = 1, \dots, n, x \in A_i\} \ .$$

On a les formules de récurrence évidentes

$$\begin{aligned} A_1 \cap \cdots \cap A_n &= (A_1 \cap \cdots \cap A_{n-1}) \cap A_n \\ A_1 \cup \cdots \cup A_n &= (A_1 \cup \cdots \cup A_{n-1}) \cup A_n \ . \end{aligned}$$

Pour finir cette section on notera que les formules de Morgan s'étendent au cas de plusieurs ensembles. C'est l'objet du lemme suivant qui regroupe les deux premiers.

**Lemme 10.3** (Formules de Morgan étendues)**.** *Soit $\Omega$ un ensemble. On a:*

*(1)* $(A_1 \cup \cdots \cup A_p) \cap B = (A_1 \cap B) \cup \cdots \cup (A_p \cap B)$ ,

*(2)* $(A_1 \cap \cdots \cap A_p) \cup B = (A_1 \cup B) \cap \cdots \cap (A_p \cup B)$

*(3)* $(A_1 \cup \cdots \cup A_p)^c = A_1^c \cap \cdots \cap A_p^c$ ,

*(4)* $(A_1 \cap \cdots \cap A_p)^c = A_1^c \cup \cdots \cup A_p^c$

*pour tous $A_1, \dots, A_p, B \subset \Omega$ sous ensembles de $\Omega$.*

*Démonstration.* On procède par récurrence sur $p$. Les Lemmes 10.1 et 10.2 montrent que ces formules sont vraies pour $p = 2$. Les amorces dans chacun des cas sont donc vérifiées. Pour ce qui est des hérédités on suppose que les propriétés sont vraies à l'ordre $p$. Pour ce qui est de (1) on écrira alors que

$$\begin{aligned} (A_1 \cup \cdots \cup A_{p+1}) \cap B &= ((A_1 \cup \cdots \cup A_p) \cup A_{p+1}) \cap B \\ &= ((A_1 \cup \cdots \cup A_p) \cap B) \cup (A_{p+1} \cap B) \\ &= (A_1 \cap B) \cup \cdots \cup (A_{p+1} \cap B) \end{aligned}$$

où la seconde ligne s'obtient avec le Lemma 10.1 et la troisième s'obtient par hypothèse de récurrence (et les formules de récurrences ci-dessus). Donc la propriété (1) à l'ordre $p$ implique la propriété (1) à l'ordre $p + 1$. Le même type de raisonnement pour les propriétés (2), (3) et (4) terminent les récurrences. D'où le lemme. □

## 11. Mathématisation de la notion de probabilité

La mathématisation de ce qui a été dit à la Section 9 est donnée par la définition qui suit. Elle reprend ce qui a été dit dans cette même Section 9, à ceci près que la formule des probabilités utilisée dans cette section correspondait au cas d'équiprobabilité, or on pourrait en imaginer d'autres. D'où une définition plus abstraite de la notion de probabilité. Si $\Omega$ est un ensemble, on note $\mathcal{P}(\Omega)$ l'ensemble de tous les sous ensembles de $\Omega$, ensemble vide $\emptyset$ compris.

**Définition 11.1** (Espace de probabilité)**.** *Soient $\Omega$ un ensemble fini et $\mathcal{P}(\Omega)$ l'ensemble des sous ensembles de $\Omega$. On appelle probabilité sur $(\Omega, \mathcal{P}(\Omega))$ toute application $P : \mathcal{P}(\Omega) \to [0, 1]$ qui vérifie les deux points suivants:*

*(1)* $P(\Omega) = 1$,

*(2)* $\forall A, B \in \mathcal{P}(\Omega)$, $A \cap B = \emptyset \ \Rightarrow \ P(A \cup B) = P(A) + P(B)$.

*Le triplet $(\Omega, \mathcal{P}(\Omega), P)$ est appelé espace de probabilité fini. On dit que $\Omega$ est l'univers des possibles (ou encore l'ensemble fondamental). Les événements sont*

*les éléments $A \in \mathcal{P}(\Omega)$, donc les sous ensembles de $\Omega$. Le réel $P(A)$ est appelé probabilité de l'événement $A$.*

L'ensemble $\Omega$ étant fini, $\mathcal{P}(\Omega)$ est lui aussi fini. Fabriquer un sous ensemble de $\Omega$ consiste à sélectionner des éléments de $\Omega$. Disons OUI si l'élément va dans le sous ensemble, et NON sinon. On peut donc assimiler les sous ensembles de $\Omega$ aux applications de $\Omega$ dans l'ensemble à deux éléments $\{\text{OUI}, \text{NON}\}$. Donc, cf. Section 4, on a $\sharp\mathcal{P}(\Omega) = 2^{\sharp\Omega}$.

**Définition 11.2.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Deux événements $A$ et $B$ tels que $A \cap B = \emptyset$ sont aussi dits incompatibles.*

## 12. Propriétés 1

On étudie les premières propriétés qui découlent de la Définition 11.1. Une propriété facile à obtenir est donnée par la proposition suivante.

**Proposition 12.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. On a toujours $P(\emptyset) = 0$.*

*Démonstration.* On a $\Omega = \Omega \cup \emptyset$ et $\Omega \cap \emptyset = \emptyset$. On doit donc avoir $P(\Omega) = P(\Omega) + P(\emptyset)$, ce qui impose $P(\emptyset) = 0$. La proposition est démontrée. □

Des sous ensembles $A_1, \ldots, A_p$ d'un ensemble $\Omega$ sont dits deux à deux disjoints si $A_i \cap A_j = \emptyset$ pour tous $i \neq j$.

**Proposition 12.2** (Additivité finie)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soient $A_1, \ldots, A_p$ ($p \geq 2$) des événements deux à deux disjoints. Alors*

$$P(A_1 \cup \cdots \cup A_p) = \sum_{i=1}^{p} P(A_i) \ .$$

*La probabilité d'une union d'événements deux à deux disjoints est la somme des probabilités des événements pris individuellement.*

*Démonstration.* On montre la formule de la proposition par récurrence sur $p$. La propriété dépendant de $p$ que l'on souhaite démontrer est la propriété:

$$\mathcal{P}(p) \ : \forall A_1, \ldots, A_p \subset \Omega, A_i \cap A_j = \emptyset \text{ pour tous } i \neq j$$
$$\Rightarrow \ P(A_1 \cup \cdots \cup A_p) = \sum_{i=1}^{p} P(A_i) \ .$$

Si $p = 2$ il s'agit de la propriété (2) de la Définition 11.1. L'amorce de la récurrence est vérifiée. On s'attaque maintenant à l'hérédité. On suppose que pour un $p$ quelconque fixé, $\mathcal{P}(p)$ est vraie. On veut montrer $\mathcal{P}(p+1)$. On considère donc $A_1, \ldots, A_{p+1}$ $(p+1)$-sous ensembles deux à deux disjoints. On pose

$$A = A_1 \cup \cdots \cup A_p \ .$$

Alors $A \cap A_{p+1} = \emptyset$. On peut le voir en notant qu'il ne peut y avoir d'élément de $A$ qui soit dans $A_{p+1}$ puisqu'un élément de $A$ est forcément dans l'un des $A_1, \ldots, A_p$, ou alors comme une conséquence dirècte des formules de Morgan étendues du Lemme 10.3. D'après la propriété (2) de la Définition 11.1,

$$P(A \cup A_{p+1}) = P(A) + P(A_{p+1}) \ . \tag{12.1}$$

Par hypothèse de récurrence,

$$P(A) = P(A_1) + \cdots + P(A_p) \ . \tag{12.2}$$

De (12.1) et (12.2) on tire facilement que

$$P(A_1 \cup \cdots \cup A_{p+1}) = \sum_{i=1}^{p+1} P(A_i) \ ,$$

et donc que $\mathcal{P}(p+1)$ est vraie elle aussi, ce qui achève la récurrence. La proposition est démontrée. □

Les propriétés du théorème suivant seront très souvent utilisées.

**Théorème 12.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Alors:*

*(1)* $P(A^c) = 1 - P(A)$ ,

*(2) Si* $A \subset B$, $P(A) \leq P(B)$ ,

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

*pour tous événements* $A, B \in \mathcal{P}(\Omega)$.

*Démonstration.* • On démontre (1). Soit $A \in \mathcal{P}(\Omega)$. On a $A \cap A^c = \emptyset$ et $A \cup A^c = \Omega$. Les deux propriétés de la Définition 11.1 permettent alors d'écrire que

$$P(A \cup A^c) = P(A) + P(A^c)$$

et que $P(A \cup A^c) = 1$. D'où la relation $P(A^c) = 1 - P(A)$.

• On démontre (2). Soient $A, B \in \mathcal{P}(\Omega)$ avec $A \subset B$. On a $\Omega = A \cup A^c$ et $B = B \cap \Omega$ puisque $B \subset \Omega$. Les lois de Morgan du Lemme 10.1 nous permettent d'écrire que $B = (B \cap A) \cup (B \cap A^c)$. On a $(B \cap A) \cap (B \cap A^c) = B \cap (A \cap A^c) = \emptyset$, et la propriété (2) de la Définition 11.1 permet d'écrire que

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

de sorte que $P(B) \geq P(A \cap B)$ puisqu'une probabilité étant à valeurs dans $[0, 1]$ est toujours positive ou nulle. Dans notre cas de figure $A \subset B$. Donc $A \cap B = A$ et $P(B) \geq P(A)$. D'où (2).

• On démontre (3). Soient $A, B \in \mathcal{P}(\Omega)$. Là encore, comme précédemment, on peut écrire que $B = (B \cap A) \cup (B \cap A^c)$. Donc

$$\begin{aligned} A \cup B &= A \cup (B \cap A) \cup (B \cap A^c) \\ &= A \cup (B \cap A^c) \end{aligned}$$

puisque $B \cap A \subset A$. Or $A \cap (B \cap A^c) = \emptyset$ puisque $B \cap A^c \subset A^c$. Donc, d'après la propriété (2) de la Définition 11.1 on peut écrire que

$$P(A \cup B) = P(A) + P(B \cap A^c) \ . \tag{12.3}$$

On a aussi $(B \cap A) \cap (B \cap A^c) = \emptyset$ puisque $B \cap A \subset A$ et $B \cap A^c \subset A^c$. Et donc, avec la propriété (2) de la Définition 11.1,

$$P(A \cap B) + P(B \cap A^c) = P(B) \tag{12.4}$$

puisque $B = (B \cap A) \cup (B \cap A^c)$. De (12.4) on tire que

$$P(B \cap A^c) = P(B) - P(A \cap B) \ , \tag{12.5}$$

et il suit de (12.3) et (12.5) que (3) est vraie. D'où le Théorème. □

## 13. Équiprobabilité

La caractérisation générale d'une probabilité est donnée par le théorème suivant qui dit essentiellement qu'une probabilité $P$ sur $\Omega$ est caractérisée par la connaissance de tous les $P(\{x\})$ pour $x \in \Omega$.

**Théorème 13.1.** *Soit $\Omega$ un ensemble quelconque fini (non vide). On note $n = \sharp\Omega$ et on écrit que $\Omega = \{\omega_1, \dots, \omega_n\}$. La donnée d'une probabilité $P : \mathcal{P}(\Omega) \to [0,1]$ est équivalente à la donnée de $n$ réels $p_i \in [0,1]$, vérifiant que $\sum_{i=1}^n p_i = 1$, via les formules $P(\{\omega_i\}) = p_i$ pour tout $i = 1, \dots, n$ et*

$$P(A) = \sum_{i \in X_A} p_i$$

*pour tout événement $A \in \mathcal{P}(\Omega)$, où, dans la formule ci-dessus, $X_A = \{i \ / \ \omega_i \in A\}$. En d'autres termes, une probabilité est entièrement caractérisée par la connaissance des $p_i = P(\{\omega_i\})$, qui vérifient forcément que $\sum_{i=1}^n p_i = 1$. On parle d'événements élémentaires pour les $\{\omega_i\}$ et de probabilités élémentaires pour les $p_i$. Les $p_i$ forment la loi de la probabilité $P$.*

*Démonstration.* Supposons que l'on ait une probabilité $P$ sur $\Omega$, et donc un espace de probabilité fini $(\Omega, \mathcal{P}(\Omega), P)$. Posons $p_i = P(\{\omega_i\})$. Comme les ensembles $\{\omega_i\}$ sont deux à deux disjoints pour $i = 1, \dots, n$, il suit de la propriété (1) de la Définition 11.1 et de la Proposition 12.2 que

$$1 = P(\Omega) = \sum_{i=1}^n P(\{\omega_i\}) = \sum_{i=1}^n p_i$$

puisque $\Omega = \{\omega_1\} \cup \dots \cup \{\omega_n\}$. Comme on a aussi $A = \cup_{i \in X_A} \{\omega_i\}$, on peut là encore écrire avec la Proposition 12.2 que $P(A) = \sum_{i \in X_A} p_i$. Réciproquement supposons que l'on se soit donné des $p_i$ tels que $\sum_{i=1}^n p_i = 1$ et que l'on pose

$$P(A) = \sum_{i \in X_A} p_i$$

pour tout événement $A$. Comme $X_\Omega = \{1, \dots, n\}$, et comme $\sum_{i=1}^n p_i = 1$, on voit à partir de cette formule que $P(\Omega) = 1$. La propriété (1) de la Définition 11.1 est ainsi vérifiée. De même, si $A \cap B = \emptyset$ alors clairement $X_{A \cup B} = X_A \cup X_B$ et $X_A \cap X_B = \emptyset$. Donc

$$P(A \cup B) = \sum_{i \in X_{A \cup B}} p_i = \sum_{i \in X_A} p_i + \sum_{i \in X_B} p_i = P(A) + P(B)$$

et la propriété (2) de la Définition 11.1 est elle aussi vérifiée. D'où le théorème. $\square$

Le cas d'équiprobabilité correspond au cas où les $\omega_i \in \Omega$ ont tous la même chance de se produire. Cela donne lieu à la définition formelle suivante.

**Définition 13.1** (Équiprobabilité)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Il y a équiprobabilité lorsque les probabilités des singletons de $\Omega$ (les événements élémentaires) sont toutes égales.*

Le théorème qui suit est une conséquence assez immédiate du Théorème 13.1. Il affirme l'existence et l'unicité des probabilités éuiprobables sur un ensemble fini donné.

**Théorème 13.2.** *Soit $\Omega$ un ensemble quelconque fini (non vide). Il existe une et une seule probabilité $P$ sur $\Omega$ qui soit équiprobable. Elle est caractérisée par $P(\{x\}) = 1/\sharp\Omega$ pour tout $x \in \Omega$ et on a que*

$$P(A) = \frac{\sharp A}{\sharp\Omega} = \frac{\textit{Nombre de cas favorables}}{\textit{Nombre de cas possibles}}$$

*pour tous les événements $A \in \mathcal{P}(\Omega)$.*

*Démonstration.* D'après le Théorème 13.1, une probabilité est entièrement caractérisée par la donnée des $P(x) = P(\{x\})$ pour $x \in \Omega$. S'il y a équiprobabilité, par défintion, $P(x) = P(y)$ pour tous $x, y \in \Omega$. Et comme $\sum_{x\in\Omega} P(x) = 1$ c'est que $P(x) = 1/\sharp\Omega$ pour tout $x \in \Omega$. Ensuite, toujours d'après le Théorème 13.1,

$$P(A) = \sum_{x\in A} P(x) = \frac{\sharp A}{\sharp\Omega} \ .$$

Il y a donc unicité et les formules du Théorème sont vérifiées. Pour l'existence, on pose $P(A) = \frac{\sharp A}{\sharp\Omega}$. On vérifie facilement à partir de cette formule que la propriété (1) de la Définition 11.1 est vraie. De même, si $A, B \in \mathcal{P}(\Omega)$ et $A \cap B = \emptyset$, alors $\sharp(A \cup B) = \sharp A + \sharp B$ et donc $P(A \cup B) = P(A) + P(B)$. Donc la propriété (2) de la Définition 11.1 est vraie elle aussi. Le théorème est démontré. □

## 14. Quelques exercices 2

**EXERCICE 7:** On lance deux fois un dé. On suppose que le dé n'est pas pipé. Quelle est la probabilité d'obtenir une somme supérieure ou égale à 10 ? On suppose maintenant que le dé est pipé et que $P(1) = P(3) = P(4) = P(5) = \frac{1}{8}$ tandis que $P(2) = P(6) = \frac{1}{4}$. On admet que $P(\{(i, j)\}) = P(i)P(j)$, à savoir que la probabilité d'obtenir $i$ au premier lancé puis $j$ au second vaut $P(i)P(j)$ (les deux lancers sont indépendants, cf. Section 20). Quelle est maintenant la probabilité d'obtenir une somme supérieure ou égale à 10 ?

CORRECTION: Lorsqu'un dé n'est pas pipé il y a équiprobabilité. Comme on lance deux fois un même dé l'ordre d'apparition compte. Si $\Omega$ est l'univers des possibles,

$$\Omega = \{(i, j) \ / \ i, j = 1, \ldots, 6\} \ .$$

On a donc (cf. Section 4) $\sharp\Omega = 6^2 = 36$. Les sommes supérieures ou égales à 10 sont réalisées par les sorties $(4, 6)$, $(5, 5)$, $(5, 6)$, $(6, 4)$, $(6, 5)$ et $(6, 6)$. L'événement $A$ des tirages donnant une somme supérieure ou égale à 10 est donc donné par

$$A = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\} \ .$$

On a $\sharp A = 6$. Donc, dans le premier cas, $P(A) = \frac{6}{36} = \frac{1}{6}$. Dans le second cas, on pourra déjà commencer par vérifier qu'il n'y a pas d'erreur d'énoncé et que l'on a bien que $\sum_{i=1}^{6} P(i) = 1$. Ensuite on écrira (cf. Théorème 13.1) que

$$\begin{aligned} P(A) &= 2P(\{(4, 6\}) + P(\{(5, 5\}) + 2P(\{(5, 6\}) + P(\{(6, 6)\}) \\ &= 2P(4)P(6) + P(5)^2 + 2P(5)P(6) + P(6)^2 \\ &= \frac{1}{16} + \frac{1}{64} + \frac{1}{16} + \frac{1}{16} \\ &= \frac{13}{64} \ . \end{aligned}$$

On a $\frac{13}{64} > \frac{1}{6}$, le dé pipé est plus intéressant.

**EXERCICE 8:** Une urne contient 12 boules numérotées de 1 à 12. On tire une boule au hasard. On considère les deux événements $A$ et $B$ donnés par $A =$ "la boule tirée a un nombre pair", $B =$ "la boule tirée a un nombre divisible par 3". Calculer $P(A)$, $P(B)$, $P(A \cap B)$. Refaire les calculs si l'urne contient 13 boules. Dans quel cas a-t-on la relation $P(A \cap B) = P(A)P(B)$ (relation qui caractérise l'indépendance, cf Section 20) ?

CORRECTION: On peut considérer que les boules ne sont pas truquées (sinon ce serait dit) et donc qu'il y a une situation d'équiprobabilité. Dans le premier cas on a $\Omega = \{1, 2, \ldots, 12\}$, $A = \{2, 4, 6, 8, 10, 12\}$, $B = \{3, 6, 9, 12\}$ et $A \cap B = \{6, 12\}$. En particulier, $\sharp\Omega = 12$, $\sharp A = 6$, $\sharp B = 4$ et $\sharp(A \cap B) = 2$. Donc (équiprobabilité):

$$P(A) = \frac{6}{12} = \frac{1}{2}\ ,\ P(B) = \frac{4}{12} = \frac{1}{3}\ ,\ P(A \cap B) = \frac{2}{12} = \frac{1}{6}\ .$$

Dans le second cas on a $\Omega = \{1, 2, \ldots, 13\}$, $A = \{2, 4, 6, 8, 10, 12\}$, $B = \{3, 6, 9, 12\}$ et $A \cap B = \{6, 12\}$. En particulier, $\sharp\Omega = 13$, $\sharp A = 6$, $\sharp B = 4$ et $\sharp(A \cap B) = 2$. Donc (équiprobabilité) Donc (toujours par équiprobabilité):

$$P(A) = \frac{6}{13}\ ,\ P(B) = \frac{4}{13}\ ,\ P(A \cap B) = \frac{2}{13}\ .$$

Dans le premier cas $P(A)P(B) = P(A \cap B)$, ce qui n'est plus vrai dans le second car $24 \neq 2 \times 13$. Dans le langage de la Section 20, $A$ et $B$ sont indépendants lorsqu'il y a 12 boules dans l'urne, mais ne le sont plus lorsqu'il y a 13 boules dans l'urne.

**EXERCICE 9$^\star$:** Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soient $A, B \in \mathcal{P}(\Omega)$ deux événements. On sait que $P(A \cup B) = \frac{7}{8}$, $P(A \cap B) = \frac{1}{4}$ et $P(A) = \frac{3}{8}$. Calculer $P(B)$, $P(A \cap B^c)$ et $P(A^c \cap B)$.

CORRECTION: On a la formule $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (cf. Théorème 12.1). Donc

$$\frac{7}{8} = \frac{3}{8} + P(B) - \frac{1}{4}$$

et on trouve que $P(B) = \frac{3}{4}$. On sait que $P(A^c) = 1 - P(A)$ et que $P(B^c) = 1 - P(B)$ (cf. là encore le Théorème 12.1), et donc $P(A^c) = \frac{5}{8}$ tandis que $P(B^c) = \frac{1}{4}$. On a $A = (A \cap B) \cup (A \cap B^c)$ et $(A \cap B) \cap (A \cap B^c) = \emptyset$. Donc, par propriété même de la définition d'une probabilité,

$$P(A) = P(A \cap B) + P(A \cap B^c)\ .$$

On trouve donc $P(A \cap B^c) = P(A) - P(A \cap B) = \frac{1}{8}$. On a aussi que

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

et donc $P(A^c \cap B) = P(B) - P(A \cap B) = \frac{1}{2}$.

**EXERCICE 10$^\star$:** Une loterie comporte 500 billets. Deux seulement sont gagnants. Combien faut-il acheter de billets pour avoir une probabilité de gain supérieure à $0,5$ ?

CORRECTION: L'exercice ne fait sens que si celui qui vend les billets ne triche pas. On pourra donc considérer qu'acheter des billets revient à tirer les billets d'un sac contenant 500 billets. Donc que nous sommes dans une situation d'équiprobabilité. Soit $1 \leq n \leq 500$ donné. L'expérience pour $n$ donné consiste à tirer $n$ billets dans le sac. Il s'agit de tirages sans remise et où l'ordre ne compte pas. On parle donc

de combinaisons et si $\Omega$ désigne l'univers de tous les tirages de $n$ billets possibles on a $\sharp\Omega = \binom{500}{n}$ (cf. Section 4). On considère l'événement

$$A = \text{“au moins un des billets est gagnant”}.$$

On veut calculer $P(A)$. Il est plus facile ici de calculer $P(A^c)$ sachant que (cf. Théorème 12.1), quand on a l'une, on a l'autre. L'événement $A^c$ correspond à la situation où les tirages ne contiennent aucun des deux billets gagnants. On a donc $\sharp A^c = \binom{498}{n}$. Et ainsi (cf. Théorème 12.1),

$$P(A) = 1 - P(A^c) = 1 - \frac{\binom{498}{n}}{\binom{500}{n}}$$

soit encore

$$\begin{aligned} P(A) &= 1 - \frac{(500-n)\times(499-n)}{500\times 499} \\ &= \frac{n(999-n)}{249.500} \ . \end{aligned}$$

On veut trouver les $n$ tels que

$$\frac{n(999-n)}{249.500} \geq \frac{1}{2}$$

soit donc les $n$ tels que

$$n^2 - 999n + 124.750 \leq 0 \ .$$

Le discriminant de cette expression du second degré vaut

$$\Delta = 999^2 - 4\times 124.750 = 499.001$$

et les racines de l'expression sont donc

$$a = \frac{999-\sqrt{\Delta}}{2} \text{ et } b = \frac{999+\sqrt{\Delta}}{2}$$

soit encore $a \sim 146,3$ et $b \sim 852,7$. Un polynôme du second degré est du signe du coefficient du terme de degré 2 en dehors des racines, et change de signe à l'intérieur. De plus on a toujours $n \leq 500$. On veut donc $n \geq 147$.

**EXERCICE 11:** Au loto les joueurs choisissent 6 numéros sur une grille qui en comporte 49. La Française des jeux sélectionne 6 numéros. Quelle est la probabilité d'avoir 6 bons numéros ? Et 5 ? Et 4 ? Et au moins 4 bon numéros ?

CORRECTION: Nous sommes dans une situation d'équiprobabilité (on l'espère !). Il s'agit de listes non ordonnées sans remise, et donc le cardinal de l'univers des possibles $\Omega$ est $\sharp\Omega = \binom{49}{6} = 13.983.816$. Soit $A_p$ l'événement “il y a $p$ numéros gagnants”. On a $\sharp A_6 = 1$ et donc

$$P(A_6) = \frac{1}{13.983.816} \sim 0,00000007$$

de l'ordre de $10^{-7}$. Pour obtenir 5 numéros gagnants, comprendre exactement 5 numéros gagnants, il faut 5 numéros parmi les 6 gagnants et 1 numéro parmi les 43 non gagnants. On a donc $\sharp A_5 = \binom{43}{1}\binom{6}{5} = 43\times 6 = 258$ et ensuite

$$P(A_5) = \frac{258}{13.983.816} \sim 0,00002$$

de l'ordre de $10^{-5}$. Pour obtenir 4 numéros gagnants, comprendre exactement 4 numéros gagnants, il faut 4 numéros parmi les 6 gagnants et 2 numéros parmi les 43 non gagnants. On a donc $\sharp A_4 = \binom{43}{2}\binom{6}{4} = 903 \times 15 = 13.545$ et ensuite

$$P(A_4) = \frac{13.545}{13.983.816} \sim 0,0009$$

de l'ordre de $10^{-3}$. Si $A$ est l'événement "il y a au moins 4 numéros gagnants", comme les $A_i$ sont deux à deux disjoints, on a

$$\begin{aligned} \sharp A &= \sum_{i=4}^{6} \sharp A_i \\ &= 13.804 \ . \end{aligned}$$

D'où

$$P(A) = \frac{13.804}{13.983.816} \sim 0,0009 \ .$$

La probabilité d'obtenir au moins 4 numéros gagnants est de l'ordre de $10^{-3}$. En langage courant, de tous les jours, on a à peu près une chance sur 1.000 d'obtenir 4 bons numéros ou plus, et seulement une chance sur 14 millions d'obtenir les 6.

**EXERCICE 12:** [Paradoxe des anniversaires] Dans une classe de 35 élèves, quelle est la probabilité qu'au moins deux élèves fêtent leur anniversaire le même jour ?

CORRECTION: Une année est constituée de 365 jours. On considère que l'on est dans une situation d'équiprobabilité. L'univers des possibles $\Omega$ est constitué de tous les jours d'anniversaires possibles, et donc $\sharp\Omega = 365^{35}$. Soit $A$ l'événement "au moins deux élèves fêtent leur anniversaire le même jour". L'événement $A^c$ correspond à "les élèves ont tous des anniversaires des jours différents". Les élèves étant distinguables on cherche les familles ordonnées sans remise (vous pouvez aussi penser en termes d'injections de $\{1, \dots, 35\}$ dans $\{1, \dots, 365\}$). Donc $\sharp A^c = A_{365}^{35}$. D'où $P(A^c) = A_{365}^{35}/365^{35} \sim 0,19$. Et donc (cf. Théorème 12.1), $P(A) = 1 - P(A^c) \sim 0,81$.

## 15. Propriétés 2

On généralise dans cette section la formule (3) du Théorème 12.1. Le résultat est à mettre en parallèle avec la discussion à la fin de la Section 3. Il s'agit ici de donner une formule pour la probabilité de l'union quelconque d'événements.

**Théorème 15.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soient $A_1, \dots, A_n \in \mathcal{P}(\Omega)$ des événements. Alors*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \dots \cap A_{i_p}) \ , \tag{15.1}$$

*où $X_p^n = \{(i_1, \dots, i_p) \ / \ 1 \le i_1 < i_2 < \dots < i_p \le n\}$ est l'ensemble des $p$-uplets $(i_1, \dots, i_p)$ dans $\{1, \dots, n\}$ ordonnés par ordre strictement croissant. Lorsque $n = 2$ on retrouve bien sûr la formule $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Lorsque $n = 3$, on trouve la formule $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.*

*Démonstration.* On démontre la formule (15.1) par récurrence sur $n$. Le cas $n = 2$ a déjà été démontré au Théorème 12.1. L'amorce de la récurrence est donc vérifiée. On s'attaque maintenant à l'hérédité. On suppose que pour un certain $n$ donné quelconque, (15.1) a lieu pour toute famille d'événements $A_1, \dots, A_n$. On considère $A_1, \dots, A_{n+1}$ une famille quelconque de $n+1$ événements et on veut démontrer (15.1) pour cette famille. On pose

$$A = A_1 \cup \cdots \cup A_n \ .$$

Avec le Théorème 12.1 on peut écrire que

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P(A) + P(A_{n+1}) - P(A \cap A_{n+1}) \ . \tag{15.2}$$

Les formules de Morgan étendues du Lemme 10.3 donnent que

$$A \cap A_{n+1} = \bigcup_{i=1}^{n} A_i \cap A_{n+1} \tag{15.3}$$

et donc, par hypothèse de récurrence, puisqu'on a ici une union de $n$ événements $A_i \cap A_{n+1}$, on obtient à partir de (15.3) que

$$P(A \cap A_{n+1}) = \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \cdots \cap A_{i_p} \cap A_{n+1}) \ . \tag{15.4}$$

De même, par hypothèse de récurrrence,

$$P(A) = \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \cdots \cap A_{i_p}) \ . \tag{15.5}$$

Il suit de (15.2), (15.4) et (15.5) que

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \cdots \cap A_{i_p}) + P(A_{n+1}) \\ &\quad - \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \cdots \cap A_{i_p} \cap A_{n+1}) \ . \end{aligned} \tag{15.6}$$

Clairement, puisque $n+1$ est de fait le plus grand des indices,

$$X_p^{n+1} = X_p^n \cup Y_p \text{ et } X_p^n \cap Y_p = \emptyset \ , \tag{15.7}$$

où

$$Y_p = \Big\{(i_1, \dots, i_{p-1}, n+1) \ / \ 1 \le i_1 < \cdots < i_{p-1} \le n\Big\}$$

Il est alors facile de vérifier que

$$\begin{aligned} &\sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_p)\in X_p^n} P(A_{i_1} \cap \cdots \cap A_{i_p} \cap A_{n+1}) \\ &= \sum_{p=1}^{n} (-1)^{p+1} \sum_{(i_1,\dots,i_{p+1})\in Y_{p+1}} P(A_{i_1} \cap \cdots \cap A_{i_{p+1}}) \\ &= \sum_{p=2}^{n+1} (-1)^{p} \sum_{(i_1,\dots,i_p)\in Y_p} P(A_{i_1} \cap \cdots \cap A_{i_p}) \end{aligned}$$

de sorte que

$$\begin{aligned}&\sum_{p=1}^{n}(-1)^{p+1}\sum_{(i_1,\dots,i_p)\in X_p^n}P(A_{i_1}\cap\dots\cap A_{i_p}\cap A_{n+1})\\&=-\sum_{p=2}^{n+1}(-1)^{p+1}\sum_{(i_1,\dots,i_p)\in Y_p}P(A_{i_1}\cap\dots\cap A_{i_p})\ .\end{aligned}\tag{15.8}$$

A partir de (15.6), (15.7) et (15.8) on obtient que

$$\begin{aligned}P\left(\bigcup_{i=1}^{n+1}A_i\right)&=\sum_{p=2}^{n}(-1)^{p+1}\sum_{(i_1,\dots,i_p)\in X_p^{n+1}}P(A_{i_1}\cap\dots\cap A_{i_p})\\&\quad+P(A_{n+1})+\sum_{i=1}^{n}P(A_i)+(-1)^{n+2}P(A_1\cap\dots\cap A_{n+1})\\&=\sum_{p=1}^{n+1}(-1)^{p+1}\sum_{(i_1,\dots,i_p)\in X_p^{n+1}}P(A_{i_1}\cap\dots\cap A_{i_p})\ .\end{aligned}\tag{15.9}$$

qui n'est autre que la formule (15.1) à l'ordre $n+1$. La récurrence est achevée. Le théorème est démontré. □

## 16. Probabilités conditionnelles

La notion de probabilité conditionnelle est rattachée au calcul de la probabilité d'un événement $B$ sachant qu'un événement $A$ s'est produit. Par exemple, les 6 numéros gagnants du loto sont les $12, 15, 27, 34, 42, 46$. On peut se demander quel est la probabilité d'obtenir 5 bons numéros sachant, par exemple, que la première boule sortie est le 7 (événement B/A). On note A l'événement "une des boules est le 7". Si $\Omega$ est l'univers des possibles, $\sharp\Omega=\binom{49}{6}=13.983.816$, $\sharp A=\binom{48}{5}=1.712.304$ et si $B$ est l'événement "avoir 5 bons numéros", on a $\sharp(A\cap B)=\binom{6}{5}=6$. Essayons maintenant d'évaluer la probabilité de $B/A$. Comme on sait que $A$ a eu lieu, le cardinal de l'univers $\Omega'$ des possibles n'est plus que $\sharp\Omega'=\binom{48}{5}$. Et il y a $\binom{6}{5}=6$ cas de figures qui réalisent $B/A$ de sorte que

$$P(B/A)=\frac{6}{1.712.304}\ .$$

En fait c'est comme si on tirait 5 numéros parmi 48 (puisque le 7 est pris) et qu'il y avait 6 numéros gagnants. On vérifie que $P(A)P(B/A)=P(A\cap B)$ puisque d'une part $P(A)=\frac{1.712.304}{13.983.816}$ et, d'autre part, $P(A\cap B)=\frac{6}{13.983.816}$. La formule est en fait générale et elle caractérise les probabilités dites conditionnelles.

**Définition 16.1** (Probabilité conditionnelle)**.** *Soit $(\Omega,\mathcal{P}(\Omega),P)$ un espace de probabilité fini. Soit $A\in\mathcal{P}(\Omega)$ un événement de probabilité non nulle. Pour tout événement $B\in\mathcal{P}(\Omega)$, la probabilité conditionnelle de $B$ sachant $A$, notée $P(B/A)$, est définie par $P(B/A)=\frac{P(A\cap B)}{P(A)}$.*

Le résultat suivant a lieu.

**Proposition 16.1.** *Soit $(\Omega,\mathcal{P}(\Omega),P)$ un espace de probabilité fini. Soit $A\in\mathcal{P}(\Omega)$ un événement de probabilité non nulle. On note $P(\cdot/A)$ la probabilité conditionnelle sachant $A$. Alors $(\Omega,\mathcal{P}(\Omega),P(\cdot/A))$ est encore un espace de probabilité fini.*

*Démonstration.* Il faut vérifier les points (1) et (2) de la Définition 11.1. On a, par définition de la probabilité conditionnelle, $P(\Omega/A) = P(A \cap \Omega)/P(A) = P(A)/P(A) = 1$ puisque $\Omega \cap A = A$. Par ailleurs, si $B$ et $C$ sont deux événements avec $B \cap C = \emptyset$, alors $A \cap B$ et $A \cap C$ sont encore tels que $(A \cap B) \cap (A \cap C) = \emptyset$ et donc, avec les formules de Morgan du Lemma 10.1, et le point (2) de la Définition 11.1 pour $P$,

$$\begin{aligned} P((B \cup C)/A) &= \frac{P((B \cup C) \cap A)}{P(A)} \\ &= \frac{P((B \cap A) \cup (C \cap A))}{P(A)} \\ &= \frac{P(B \cap A) + P(C \cap A)}{P(A)} \\ &= P(B/A) + P(C/A) \ . \end{aligned}$$

Le point (2) de la Définition 11.1 pour $P(\cdot/A)$ est donc lui aussi vérifié. La proposition est démontrée. □

Le théorème qui suit dit comment retrouver la probabilité de $B$ connaissant les probabilités conditionnelles de $B$ sachant $A$ et de $B$ sachant $A^c$. On parle de formule des probabilités totales. Comme on veut parler à la fois de $P(\cdot/A)$ et $P(\cdot/A^c)$ il faut supposer que $P(A) > 0$ et $P(A^c) > 0$ et comme $P(A^c) = 1 - P(A)$ (cf. Théorème 12.1), cela revient à supposer que $0 < P(A) < 1$.

**Théorème 16.1** (Formule des probabilités totales)**.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini. Soit* $A \in \mathcal{P}(\Omega)$ *tel que* $0 < P(A) < 1$*. Alors*

$$P(B) = P(A)P(B/A) + P(A^c)P(B/A^c)$$

*pour tout événement* $B \in \mathcal{P}(\Omega)$*.*

*Démonstration.* On a clairement, à partir de la Définition 16.1, que $P(A)P(B/A) = P(A \cap B)$ et que $P(A^c)P(B/A^c) = P(A^c \cap B)$. Or $B = (A \cap B) \cup (A^c \cap B)$ et $(A \cap B) \cap (A^c \cap B) = \emptyset$. Donc $P(B) = P(A \cap B) + P(A^c \cap B)$ et on retrouve la formule des probabilités totales. Le théorème est démontré. □

UN EXEMPLE D'APPLICATION: LE PROBLÈME DE MONTY HALL. Le problème de Monty Hall est une histoire probabiliste librement inspirée du jeu télévisé *Let's make a deal* dont la version originale, présentée par Monty Hall, a été diffusée de 1963 à 1977 sur les chaînes américaine NBC puis ABC. Le candidat à ce jeu fait face à 3 portes. Derrière l'une d'entre elles se cache un prix, et rien derrière les deux autres. Le candidat désigne une porte mais ne l'ouvre pas. Le présentateur, qui sait quelle porte est gagnante, ouvre une des deux autres portes derrière laquelle il n'y a pas de prix. Le candidat a alors la possibilité de changer son choix. Soit garder la porte qu'il avait initialement choisie, soit prendre l'autre porte non encore ouverte. La question est: "le candidat a-t-il intérêt à changer d'avis ?" En d'autres termes la probabilité qu'il ouvre la bonne porte est-elle plus élevée s'il change d'avis ou s'il ne change pas d'avis ? On a là une belle illustration de la formule des probabilités totales. On note $A$ l'événement "le 1er choix du candidat était le bon" et $B$ l'événement "le candidat ouvre la bonne porte" (et donc remporte le prix). On a

$$P(B) = P(B/A)P(A) + P(B/A^c)P(A^c) \ .$$

Clairement $P(A) = 1/3$. Par suite $P(A^c) = 2/3$. Il y a maintenant deux cas de figure. Dans le premier cas: le candidat ne change pas d'avis. Alors $P(B/A) = 1$ tandis que $P(B/A^c) = 0$. On trouve alors $P(B) = P(A) = 1/3$. Dans le second cas: le candidat change d'avis. Alors $P(B/A) = 0$ tandis que $P(B/A^c) = 1$. On trouve alors $P(B) = P(A^c) = 2/3$. CONCLUSION: le candidat a tout intérêt à changer d'avis. □

On aborde maintenant la notion de système complet d'événements.

**Définition 16.2.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Une famille finie $A_1, \dots, A_n$ d'événements non-vides est appelée un système complet d'événements si les $A_i$ sont deux à deux disjoints ($A_i \cap A_j = \emptyset$ pour tous $i \neq j$) et si $\Omega = A_1 \cup \dots \cup A_n$. On dit encore que les $A_i$ forment une partition de $\Omega$ (une partition d'un ensemble $X$ est un ensemble de parties non vides de $X$ deux à deux disjointes et dont l'union est $X$).*

Si $A \in \mathcal{P}(\Omega)$ n'est ni l'ensemble vide $\emptyset$, ni $\Omega$ lui-même, alors $A, A^c$ est une partition de $\Omega$ à deux éléments ($n = 2$). La formule de la proposition suivante généralise du coup la formule des probabilités totales du Théorème 16.1 (prendre $A_1, \dots, A_n = A, A^c$).

**Proposition 16.2.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A_1, \dots, A_n$ un système complet d'événements tels que $P(A_i) > 0$ pour tout $i$. Alors*

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B/A_i)$$

*pour tout événement $B \in \mathcal{P}(\Omega)$. On parle de formule généralisée des probabilités totales.*

*Démonstration.* On a $B = \bigcup_{i=1}^{n} A_i \cap B$ et les $A_i \cap B$ sont deux à deux disjoints. Par suite, d'après la Proposition 12.2,

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$

et comme, d'après la définition des probabilités conditionnelles,

$$P(A_i \cap B) = P(A_i) P(B/A_i)$$

pour tout $i$, on retrouve bien la formule voulue. La proposition est démontrée. □

## 17. Formule de Bayes

La formule de Bayes cherche à déterminer $P(A_i/B)$ en fonction de $P(B/A_i)$. C'est une sorte de problème inverse. Elle est donnée par le théorème suivant, ainsi que par le corollaire qui suit dans le cas réduit d'une partition $A, A^c$.

**Théorème 17.1** (Formule de Bayes)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A_1, \dots, A_n$ un système complet d'événements tels que $P(A_i) > 0$ pour tout $i$. Alors*

$$P(A_i/B) = \frac{P(A_i) P(B/A_i)}{\sum_{j=1}^{n} P(A_j) P(B/A_j)} \tag{17.1}$$

*pour tout événement $B \in \mathcal{P}(\Omega)$ de probabilité non nul ($P(B) > 0$).*

*Démonstration.* On a $P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$ et, d'après la Proposition 16.2,

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i) \ .$$

Sachant qu'on a aussi $P(A_i)P(B/A_i) = P(A_i \cap B)$, on retrouve bien la formule (17.1). Le théorème est démontré. □

Le résultat qui suit est un corollaire immédiat du Théorème 17.1.

**Corollaire 17.1** (Formule de Bayes réduite)**.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini et soit* $A$ *un événement tel que* $0 < P(A) < 1$*. Alors*

$$P(A/B) = \frac{P(A)P(B/A)}{P(A)P(B/A) + P(A^c)P(B/A^c)} \tag{17.2}$$

*pour tout événement* $B \in \mathcal{P}(\Omega)$ *de probabilité non nul (*$P(B) > 0$*).*

*Démonstration.* Il suffit d'appliquer le Théorème 17.1 dans le cas d'une partition $A, A^c$. On retrouve alors (17.2). □

En d'autres termes, on connait $P(A/B)$ si on connait $P(A)$, $P(A^c)$, $P(B/A)$ et $P(B/A^c)$.

## 18. Quelques exercices 3

**EXERCICE 13:** Un artisan possède deux machines M1 et M2. La probabilité que M1 tombe en panne est de $0,01$ et la probabilité que M2 tombe en panne est de $0,008$. On sait aussi que la probabilité que M2 tombe en panne sachant que M1 est en panne est de $0,4$. Quelle est la probabilité que les deux machines tombent en panne en même temps ? Quelle est la probabilité d'avoir au moins une machine qui fonctionne ?

CORRECTION: Soit $A1$ l'événement "M1 tombe en panne" et $A2$ l'événement "M2 tombe en panne ". On cherche $P(A1 \cap A2)$. On a (cf. Définition 16.1) $P(A1 \cap A2) = P(A2/A1)P(A1)$. Donc $P(A1 \cap A2) = 0,01 \times 0,4 = 0,004$. On cherche maintenant $P(A1^c \cup A2^c)$ puisque $A1^c$ est l'événement "M1 fonctionne" et $A2^c$ est l'événement "M2 fonctionne". Les lois de Morgan (Lemme 10.2) donnent $A1^c \cup A2^c = (A1 \cap A2)^c$ et donc (cf. Théorème 12.1), $P((A1 \cap A2)^c) = 1 - P(A1 \cap A2) = 1 - 0,004 = 0,996$.

**EXERCICE 14:** Dans une population, 40% des individus ont les yeux bruns, 25% des individus ont les cheveux blonds, 15% des individus ont les yeux bruns et les cheveux blonds. On choisit un individu au hasard. Quelle est la probabilité des événements: $A$ "si un individu a les yeux bruns d'avoir les cheveux blonds", $B$ "si un individu a les cheveux blonds d'avoir les yeux bruns", $C$ "si un individu a les cheveux blonds de ne pas avoir les yeux bruns".

CORRECTION: Notons $M1$ l'événement "avoir les yeux bruns" et $M2$ l'événement "avoir les cheveux blonds". L'individu étant choisi au hasard on se trouve dans une situation d'équiprobabilité. On a alors, d'après l'énoncé, $P(M1) = 0,4$, $P(M2) = 0,25$ et $P(M1 \cap M2) = 0,15$. $P(A) = P(M2/M1)$. D'après la Définition 16.1, $P(M2/M1) = P(M1 \cap M2)/P(M1)$. On a donc

$$P(A) = 0,15/0,4 = 0,375 \ .$$

On a $P(B) = P(M1/M2)$ et, cf. Définition 16.1, $P(M1/M2) = P(M1 \cap M2)/P(M2)$. Donc

$$P(B) = 0,15/0,25 = 0,6 \ .$$

On a pour finir $P(C) = P(M1^c/M2)$ et, toujours d'après la Définition 16.1, $P(M1^c/M2) = P(M1^c \cap M2)/P(M2)$. Si $\Omega$ est l'univers des possibles, l'espace donc de tous les individus, on a $M1 \cup M1^c = \Omega$ et les formules de Morgan (cf. Lemme 10.1) permettent d'écrire que $M2 = (M1 \cap M2) \cup (M1^c \cap M2)$. On a $(M1 \cap M2) \cap (M1^c \cap M2) = \emptyset$ et ainsi

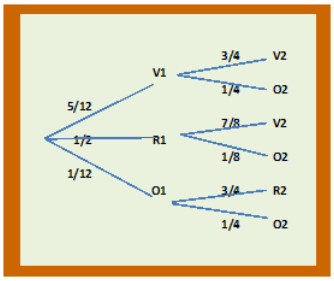
$$P(M2) = P(M1 \cap M2) + P(M1^c \cap M2) \ .$$

Donc $P(M1^c \cap M2) = 0,25 - 0,15 = 0,1$ et

$$P(C) = P(M1^c/M2) = 0,1/0,25 = 0,4 \ .$$

## 19. Formule des probabilités composées

La formule des probabilités composées correspond à la situation du calcul de probabilités à partir d'un arbre de probabilité. Un arbre de probabilités correspond à une situation où des événement se succèdent et où l'on raisonne d'étape en étape avec, du coup, des probabilités conditionnelles. Prenons la situation suivante: Un automobiliste rencontre successivement deux feux tricolores, le second étant réglé en fonction du premier. L'arbre de probabilité est donné ci-dessous. V1 signifie "le premier feu est vert", O2 signifie "le second feu est orange", etc. Les probabilités sont données sur les branches. A chaque sommet la somme des probabilités vaut 1.




Le fait que tous les cas de figures ne se rencontrent pas au second feu pour un premier feu donné illustre seulement le fait que le second feu est réglé en fonction du premier. Si par exemple on rencontre un premier feu orange alors on ne rencontrera pas de second feu vert. La probabilité $p$ que l'automobiliste rencontre les deux feux au vert est alors

$$p = \frac{5}{12} \times \frac{3}{4} = \frac{5}{16}$$

(première branche) et la probabilité $p'$ que l'automobiliste rencontre le deuxième feu au vert est

$$p' = \frac{5}{12} \times \frac{3}{4} + \frac{1}{2} \times \frac{7}{8} = \frac{5}{16} + \frac{7}{16} = \frac{3}{4}$$

(première branche plus troisième branche). Il y a addition par incompatibilité. Les branches de l'arbre donnent juste les probabilités conditionnelles et on applique à chaque fois la formule $P(A \cap B) = P(A)P(B/A)$. Par exemple, $P(V1 \cap V2) = P(V1)P(V2/V1) = \frac{5}{12} \times \frac{3}{4} = \frac{5}{16}$. La formule générale qui englobe ce problème est la formule dite des probabilités composées.

**Théorème 19.1** (Formule des probabilités composées)**.** *Soient* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini et* $A_1, \dots, A_n$ *des événements tels que* $P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$*. Alors*

$$\begin{aligned} &P(A_1 \cap \cdots \cap A_n) \\ &= P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap \cdots \cap A_{n-1}) \ . \end{aligned} \tag{19.1}$$

*Si* $n = 2$ *on retrouve la formule* $P(A \cap B) = P(A)P(B/A)$*. Si* $n = 3$ *on trouve la formule* $P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B)$*.*

On pourra remarquer qu'en vertue de la propriété (2) du Théorème 12.1, comme $P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$, donc en fait comme $P(A_1 \cap \cdots \cap A_{n-1} > 0$, on a forcément que $P(A_1 \cap \cdots \cap A_p) > 0$ pour tout $p = 1, \dots, n-1$.

*Démonstration.* On démontre cette formule par récurrence sur $n$. Si $n = 2$ il s'agit tout juste de la Définition 16.1. L'amorce est vérifiée. Pour démontrer l'hérédité on suppose que pour un $n$ donné quelconque, (19.1) est vraie pour tous les événements $A_1, \dots, A_n$ tels que $P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$. On considère $n+1$ événements $A_1, \dots, A_{n+1}$ tels que $P(A_1 \cap \cdots \cap A_n) \neq 0$. Soit $A = A_1 \cap \cdots \cap A_n$. On a

$$\begin{aligned} P(A_1 \cap \cdots \cap A_{n+1}) &= P(A \cap A_{n+1}) \\ &= P(A_{n+1}/A)P(A) \end{aligned}$$

et par hypothèse de récurrence on récupère que

$$\begin{aligned} &P(A_1 \cap \cdots \cap A_{n+1}) \\ &= P(A_{n+1}/A)P(A) \\ &= P(A_1)P(A_2/A_1) \dots P(A_n/A_1 \cap \cdots \cap A_{n-1})P(A_{n+1}/A) \\ &= P(A_1)P(A_2/A_1) \dots P(A_n/A_1 \cap \cdots \cap A_{n-1})P(A_{n+1}/A_1 \cap \cdots \cap A_n) \end{aligned}$$

qui n'est rien d'autre que la formule (19.1) pour les événements $A_1, \dots, A_{n+1}$. L'hérédité est donc vérifiée. Le théorème est démontré. □

**EXERCICE 15:** Une urne contient initialement 7 boules noires et 3 boules blanches. On tire successivement 3 boules : si on tire une noire, on l'enlève, si on tire une blanche, on la retire, et on ajoute une noire à la place. Quelle est la probabilité de tirer 3 boules blanches à la suite?

CORRECTION: Soit $B_i$ l'événement "la $i$ème boule tirée est blanche", $i = 1, 2, 3$. On veut calculer $P(B_1 \cap B_2 \cap B_3)$. La formule (19.1) donne que

$$P(B_1 \cap B_2 \cap B_3) = P(B_1)P(B_2/B_1)P(B_3/B_1 \cap B_2) \ .$$

Le tirage des boules correspond à une situation d'équiprobabilité. On a $P(B_1) = \frac{3}{10}$. Par ailleurs, $P(B_2/B_1) = \frac{2}{10} = \frac{1}{5}$ puisque si la première boule tirée est blanche l'urne contient alors 8 boules noires et deux blanches. Enfin $P(B_3/B_1 \cap B_2) = \frac{1}{10}$ puisque si les deux premières boules tirées sont blanches alors l'urne contient 9 boules noires et une boule blanche pour le 3ème tirage. D'où le résultat pour la probabilité demandée $P(B_1 \cap B_2 \cap B_3) = \frac{6}{1000} = \frac{3}{500}$.

## 20. Indépendance

Dans de nombreuses situations concrètes, la réalisation d'un événement $A$ n'influe en rien sur la réalisation ou pas d'un événement $B$. Donc $P(B/A) = P(B)$. Dans ce cas on dit que les événements $A$ et $B$ sont indépendants. La définition formelle de l'indépendance de deux événements est donnée ci-dessous.

**Définition 20.1** (Evénements indépendants)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Deux événements $A$ et $B$ sont dits indépendants si $P(A \cap B) = P(A)P(B)$*

L'indépendance peut être plus difficile qu'on ne l'imagine à deviner. Revenons à l'exercice 8. Une urne contient 12 boules numérotées de 1 à 12. On tire une boule au hasard. On considère les deux événements $A$ et $B$ donnés par $A =$ "la boule tirée a un nombre pair", $B =$ "la boule tirée a un nombre divisible par 3". Ces événements, on l'a vu, sont indépendants. Si par contre l'urne contient 13 boules numérotées de 1 à 13 les événements $A$ et $B$ cessent d'être indépendants (cf. là encore la correction de l'exercice).

Attention à ne pas confondre indépendance et incompatibilité. Si deux événements sont incompatibles ils influent l'un sur l'autre puisque l'un ne peut se produire si l'autre a eu lieu. Sauf si $P(A) = 0$ ou $P(B) = 0$, deux événements incompatibles ne sont jamais indépendants puisque pour deux événements incompatibles $P(A \cap B) = 0$ (cf. Proposition 12.1).

**Théorème 20.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A$ et $B$ deux événements. Si $A$ et $B$ sont indépendants, alors $A^c$ et $B$ sont indépendants, $A$ et $B^c$ sont indépendants et $A^c$ et $B^c$ sont indépendants.*

*Démonstration.* Par hypothèse $P(A \cap B) = P(A)P(B)$. Puisque $\Omega = A \cup A^c$, les formules de Morgan du Lemme 10.1 donnent que $B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$ et comme $A \cap A^c = \emptyset \Rightarrow (A \cap B) \cap (A^c \cap B) = \emptyset$, on peut écrire que

$$P(B) = P(A \cap B) + P(A^c \cap B) \ .$$

Donc

$$P(A^c \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(A^c)P(B)$$

puisque $P(A^c) = 1 - P(A)$ (cf. Théorème 12.1). Les événements $A^c$ et $B$ sont donc bien indépendants. Par symétrie en $A$ et $B$ les événements $A$ et $B^c$ sont tout aussi indépendants. On vient de montrer que si $A$ et $B$ sont indépendants alors $A^c$ et $B$ le sont aussi. On ré-applique cela à l'indépendance de $B$ et $A^c$. Alors $B^c$ et $A^c$ sont indépendants. Et donc $A^c$ et $B^c$ sont eux encore aussi indépendants. Le théorème est démontré. □

On aborde pour finir l'indépendance d'une famille d'événements.

**Définition 20.2** (Famille d'événements indépendants)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Une famille $A_1, \ldots, A_n$ d'événements est dite indépendante (ou mutuellement indépendante) si pour tout $1 \leq p \leq n$ et toute sous famille $A_{i_1}, \ldots, A_{i_p}$ de $A_1, \ldots, A_n$, $P(A_{i_1} \cap \cdots \cap A_{i_p}) = P(A_{i_1}) \times \cdots \times P(A_{i_p})$*

Au lieu de dire que la famille $A_1, \ldots, A_n$ d'événements est indépendante (ou mutuellement indépendante) on dira plus simplement que les événements $A_1, \ldots, A_n$ sont indépendants (ou mutuellement indépendants). On fera attention à ne pas

confondre la notion d'événements mutuellement indépendants que nous traitons ici et la notion d'événements deux à deux indépendants (qui correspondrait à la condition $A_i$ et $A_j$ sont indépendants pour tous $i \neq j$.) La notion d'événements mutuellement indépendants est plus exigente (mutuellement indépendants $\Rightarrow$ deux à deux indépendants). La généralisation du Théorème 20.1 au cas de figure de plusieurs événements mutuellement indépendants est donnée par le résultat suivant.

**Théorème 20.2.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A_1$,...,$A_n$ une famille indépendante d'événements. Notons $B_i$ pour $A_i$ ou $A_i^c$ au choix. La famille d'événements $B_1, \ldots, B_n$ est alors encore une famille indépendante.*

*Démonstration.* On démontre le théorème par récurrence sur $n$. Si $n = 2$ il s'agit tout simplement du Théorème 20.1. L'amorce est vérifiée. On s'attache maintenant à démontrer l'hérédité. On suppose que pour un $n \geq 2$ donné quelconque, pour toute famille $A_1, \ldots, A_n$ d'événements mutuellement indépendants, les $B_1, \ldots, B_n$ sont eux aussi mutuellement indépendants lorsque $B_i = A_i$ ou $A_i^c$. On considère $A_1$,...,$A_{n+1}$ une famille indépendante d'événements. Il est facile de se convaincre que toute sous famille d'une famille mutuellement indépendante est encore mutuellement indépendante (tout simplement parce qu'une sous famille $Z$ d'une sous famille $Y$ d'une famille $X$ est une sous famille de $X$ puisque $Z \subset Y$ et $Y \subset X$ entraînent $Z \subset X$). On obtient donc clairement que si $1 \leq p \leq n$ alors

$$P(B_{i_1} \cap \cdots \cap B_{i_p}) = P(B_{i_1}) \times \cdots \times P(B_{i_p})$$

pour tout choix d'indices $i_1, \ldots, i_p$ dans $\{1, \ldots, n+1\}$, où $B_i = A_i$ ou $A_i^c$. Pour obtenir que les $B_1, \ldots, B_{n+1}$ sont mutuellement indépendants il suffit donc de démontrer que

$$P(B_1 \cap \cdots \cap B_{n+1}) = P(B_1) \times \cdots \times P(B_{n+1}) \tag{20.1}$$

pour tout choix des $B_i$. Supposons donc un choix des $B_i$ fait. Supposons que dans ce choix deux au moins des $B_i$ soient des $A_i$. Quitte à renuméroter, disons que $B_n = A_n$ et $B_{n+1} = A_{n+1}$. Posons $A = A_n \cap A_{n+1}$. Clairement la famille $A_1, \ldots, A_{n-1}, A$ est indépendante. Elle comporte $n$ éléments. Donc, par hypothèse de récurrence, $P(B_1 \cap \cdots \cap B_{n+1}) = P(B_1) \times \cdots \times P(B_{n-1}) \times P(A)$ et comme les événements $A_n$ et $A_{n+1}$ sont aussi clairement indépendants, $P(A) = P(A_n)P(A_{n+1})$. On obtient alors (20.1). D'un autre côté, s'il est impossible de trouver deux des $B_i$ qui soient des $A_i$, comme $n + 1 \geq 3$, c'est que deux au moins des $B_i$ sont des $A_i^c$. Quitte à renuméroter on peut supposer qu'il s'agit de $B_n = A_n^c$ et $B_{n+1} = A_{n+1}^c$. On pose $B = B_1 \cap \cdots \cap B_{n-1}$. On écrit que

$$P(B \cap A_n^c) = P(B \cap A_n^c \cap A_{n+1}^c) + P(B \cap A_n^c \cap A_{n+1})$$

puis que

$$P(B \cap A_{n+1}) = P(B \cap A_n \cap A_{n+1}) + P(B \cap A_n^c \cap A_{n+1})$$

en utilisant que $\Omega = A_{n+1} \cup A_{n+1}^c$ (et $A_{n+1} \cap A_{n+1}^c = \emptyset$) ainsi que $\Omega = A_n \cup A_n^c$ (et $A_n \cap A_n^c = \emptyset$). En vertue de l'hypothèse de récurrence, et de la remarque faite plus haut lorsque deux des $B_i$ sont des $A_i$, on a que

$$\begin{aligned}
P(B \cap A_n^c) &= P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_n^c) \\
P(B \cap A_{n+1}) &= P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_{n+1}) \\
P(B \cap A_n \cap A_{n+1}) &= P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_n) \times P(A_{n+1})
\end{aligned}$$

Comme $P(B_1 \cap \cdots \cap B_{n+1}) = P(B \cap A_n^c \cap A_{n+1}^c)$ dans le cas ici étudié, c'est que

$$\begin{aligned}
P(B_1 \cap \cdots \cap B_{n+1}) &= P(B \cap A_n^c) - P(B \cap A_n^c \cap A_{n+1}) \\
&= P(B \cap A_n^c) - P(B \cap A_{n+1}) + P(B \cap A_n \cap A_{n+1}) \\
&= P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_n^c) \\
&\quad - P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_{n+1}) \\
&\quad + P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_n) \times P(A_{n+1}) \\
&= P(B_1) \times \cdots \times P(B_{n-1}) \\
&\quad \times (P(A_n^c) - P(A_{n+1}) + P(A_n)P(A_{n+1})) \\
&= P(B_1) \times \cdots \times P(B_{n-1}) \times P(A_n^c) \times P(A_{n+1}^c)
\end{aligned}$$

puisque

$$\begin{aligned}
&P(A_n^c) - P(A_{n+1}) + P(A_n)P(A_{n+1}) \\
&= 1 - P(A_n) - P(A_{n+1}) + P(A_n)P(A_{n+1}) \\
&= (1 - P(A_n)) \times (1 - P(A_{n+1})) \\
&= P(A_n^c) \times P(A_{n+1}^c) \ .
\end{aligned}$$

Comme $A_n^c = B_n$ et $A_{n+1}^c = B_{n+1}$, on retrouve là encore (20.1). L'hérédité est ainsi vérifiée, ce qui achève la récurrence. Le théorème est démontré. □

Lorsque l'on a des éléments mutuellement indépendants $A_1, \ldots, A_n$ on peut calculer $P(A_1 \cup \cdots \cup A_n)$ en fonction des $P(A_i)$. On renvoie pour cela à l'exercice 17 ci-dessous.

## 21. Quelques exercices 4

**EXERCICE 16:** Montrer que lorsque l'on tire une carte au hasard dans un jeu de 52 cartes, les événements "tirer une couleur" (cœur, carreau, trèfle, pique) et "tirer une valeur" (deux, trois, valets, as etc.) sont indépendants.

CORRECTION: Tirer une carte au hasard est bien sur une situation d'équiprobabilité. L'univers des possibles $\Omega$ est constitué des 52 cartes. Soit $A$ l'événement on tire un carreau (vous pouvez remplacer par n'importe quelle autre couleur, cela ne change rien), et soit $B$ l'événement on tire un valet (vous pouvez remplacer par n'importe quelle autre valeur, cela ne change rien). On a $\sharp A = 52/4 = 13$ et $\sharp B = 4$. Par ailleurs $\sharp(A \cap B) = 1$ de façon évidente. On a donc $P(A \cap B) = \frac{1}{52}$ tandis que

$$P(A)P(B) = \frac{13}{52} \times \frac{4}{52} = \frac{1}{52}$$

et on trouve bien que $P(A \cap B) = P(A)P(B)$. Les événements sont indépendants.

**EXERCICE 17$^\star$:** Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A_1, \ldots, A_n$ une famille indépendante d'événements. Que vaut $P(A_1 \cup \cdots \cup A_n)$ en fonction des $P(A_i)$ ?

CORRECTION: A priori l'union n'est pas vraiment dans le langage des événements indépendants. Mais en fait, si. Les formules de Morgan étendues du Lemme 10.3 donnent que $(A_1 \cup \cdots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c$ et l'on sait, avec le Théorème 20.2,

que les $A_1^c, \dots, A_n^c$ sont mutuellement indépendants. Donc (cf. Théorème 12.1) on peut écrire que

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= 1 - P((A_1 \cup \cdots \cup A_n)^c) \\ &= 1 - P(A_1^c) \times \cdots \times P(A_n^c) \\ &= 1 - (1 - P(A_1)) \times \cdots \times (1 - P(A_n)) \end{aligned}$$

ce qui répond a la question posée.

**EXERCICE 18:** Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini avec $P$ équiprobable. On suppose que $\sharp\Omega$ est un nombre premier. Montrer que si $A$ et $B$ sont deux événements indépendants, alors forcément l'un des deux est soit $\emptyset$ soit $\Omega$ tout entier.

CORRECTION: On a $\Omega \neq \emptyset$ (puisque $\sharp\Omega$ est premier). Soient $A$ et $B$ deux événements indépendants. On note $n = \sharp\Omega$, $p = \sharp A$ et $q = \sharp B$. On suppose $p \geq 1$ et $q \geq 1$ (sinon $A = \emptyset$ ou $B = \emptyset$). On a alors

$$P(A \cap B) = P(A)P(B) = \frac{pq}{n^2}$$

par équiprobabilité et indépendance. Et

$$P(A \cap B) = \frac{\sharp(A \cap B)}{n}$$

toujours par équiprobabilité. Donc

$$\sharp(A \cap B) = \frac{pq}{n}$$

et comme $\sharp(A \cap B) \in \mathbb{N}$ on obtient soit que $\sharp(A \cap B) = 0$, et donc que $pq = 0$, ce qui est impossible puisqu'on a supposé $p \geq 1$ et $q \geq 1$, soit que $n$ divise $pq$. Mais si $n$ divise $pq$, comme $n$ est premier c'est que $n$ divise $p$ et/ou $n$ divise $q$ ($n$ doit forcément apparaître dans la décomposition en nombres premiers de $p$ ou de $q$). Or $p \leq n$ et $q \leq n$. Donc forcément $p = n$ et/ou $q = n$ et $A = \Omega$ et/ou $B = \Omega$. On a donc bien montré que dans une situation d'équiprobabilité, si le cardinal de l'univers des possibles est un nombre premier, alors deux événements non triviaux (à savoir différents de $\emptyset$ et $\Omega$) ne peuvent être indépendants, ce qu'il fallait démontrer.

**EXERCICE 19:** On lance deux dés non pipés, un noir et un blanc. On considère les événements: $A$ "le chiffre du dé noir est pair", $B$ "le chiffre du dé blanc est pair", $C$ "les chiffres des deux dés ont même parité". Montrer que les événements $A, B, C$ sont deux à deux indépendants mais pas mutuellement indépendants.

CORRECTION: Les dés ne sont pas pipés et il y a donc équiprobabilité. Les "tirages" que l'on obtient sont du type N3B5, N1B2 etc. signifiant par là que dans un lancer on obtient 3 avec le dé noir et 5 avec le dé blanc, ou que dans un autre lancer on obtient 1 avec le dé noir et 2 avec le dé blanc. L'univers des possibles $\Omega$ est constitué de ces éléments. Comme on distingue les dés on peut considérer que les lancers sont séparés dans le temps: on lance d'abord le dé noir, puis on lance ensuite le dé blanc. Du coup l'univers des possibles $\Omega$ est constitué des résultats $(3, 5)$, $(1, 2)$ etc. de ces lancers et $\sharp\Omega = 6^2 = 36$ (familles ordonnées avec remise, cf. Théorème 4.2). On a $\sharp A = 3 \times 6 = 18$, $\sharp B = 6 \times 3 = 18$ et $\sharp(A \cap B) = 3 \times 3 = 9$. Pour trouver $\sharp C$ on peut raisonner de la façon suivante: il y a 3 choix pour que le premier dé lancé (le noir) soit pair et 3 choix pour qu'il soit impair. Si le premier lancé est pair, il y

a 3 choix pour que le second lancé soit pair lui aussi, et même chose en imparité. Donc $\sharp C = 2 \times 3 \times 3 = 18$. On a $\sharp(A \cap C) = 3 \times 3 = 9$, $\sharp(B \cap C) = 3 \times 3 = 9$ et enfin $\sharp(A \cap B \cap C) = \sharp(A \cap B) = 9$. On en déduit

$$P(A) = \frac{1}{2}\ ,\ P(B) = \frac{1}{2}\ ,\ P(C) = \frac{1}{2}\ ,\ P(A \cap B) = \frac{1}{4}\ ,$$
$$P(A \cap C) = \frac{1}{4}\ ,\ P(B \cap C) = \frac{1}{4},\ P(A \cap B \cap C) = \frac{1}{4}.$$

Donc $P(A \cap B) = P(A)P(B)$ et $A$ et $B$ sont indépendants, $P(A \cap C) = P(A)P(C)$ et $A$ et $C$ sont indépendants, et $P(B \cap C) = P(B)P(C)$ et $B$ et $C$ sont indépendants. En d'autres termes, $A, B, C$ sont deux à deux indépendants. Par contre

$$P(A)P(B)P(C) = \frac{1}{8}$$

de sorte que $P(A \cap B \cap C) \neq P(A)P(B)P(C)$ et $A, B, C$ ne sont donc pas mutuellement indépendants.

# PARTIE III
## VARIABLES ALÉATOIRES

On aborde dans cette dernière partie la théorie rattachée aux variables aléatoires dans le cadre des probabilités discrètes finies. Les variables aléatoires suivent des lois, certaines sont très connues (comme la loi binomiale), elles ont une espérance, une variance etc.

### 22. Premières définitions

On commence tout naturellement avec la défintion importante d'une variable aléatoire.

**Définition 22.1** (Variables aléatoires)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ est une application $X : \Omega \to E$ à valeurs dans un ensemble fini $E$ (souvent un sous ensemble fini de $\mathbb{Z}$ ou de $\mathbb{R}$). La loi de $X$ est la fonction $P_X : \mathcal{P}(E) \to [0,1]$ définie par*

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

*pour tout $A \in \mathcal{P}(E)$, où la notation $P(X \in A)$ signifie en fait $P(\{X \in A\})$ et où $\{X \in A\} \in \mathcal{P}(\Omega)$ est donné par $\{X \in A\} = \{\omega \in \Omega \ / \ X(\omega) \in A\}$.*

Autant nous avons pu éviter cette notation jusqu'ici, autant elle devient vraiment utile dans le contexte des variables aléatoires. Etant donnés $p \leq q$ deux entiers, on note $[\![p, q]\!]$ le sous ensemble de $\mathbb{N}$ constitué de tous les entiers de $p$ (compris) à $q$ (compris), et donc

$$[\![p, q]\!] = \{p, p+1, \ldots, q\} \ .$$

Par exemple $[\![1, 6]\!] = \{1, 2, 3, 4, 5, 6\}$. Revenons maintenant à l'exemple de Galilée. On lance donc trois dés: l'un, puis un autre, puis un dernier (de manière à les distinguer). L'univers des possibles est alors $\Omega = [\![1, 6]\!]^3$. Si on s'intéresse à la somme des trois dés, la variable aléatoire $X$ qui correspond à cette somme est l'application $X : \Omega \to [\![3, 18]\!]$ qui à $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ associe $X(\omega) = \omega_1 + \omega_2 + \omega_3$. Ici, donc, $E = [\![3, 18]\!]$.

Le théorème qui suit est fondamental dans la compréhension de $P_X$. La loi $P_X$ est effectivement une loi de probabilité. Elle l'est sur $(E, \mathcal{P}(E))$.

**Théorème 22.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$. Soit $P_X$ la loi de $X$. Alors $P_X$ est une loi de probabilité sur $(E, \mathcal{P}(E))$.*

*Démonstration.* Il faut démontrer que $P_X$ vérifie les deux propriétés de la Définition 11.1. Clairement, puisque $X$ est à valeurs dans $E$, $\{X \in E\} = \Omega$ et donc $P_X(E) = P(\Omega) = 1$. Considérons maintenant $A, B \in \mathcal{P}(E)$ tels que $A \cap B = \emptyset$. Alors, clairement

$$\{X \in A\} \cap \{X \in B\} = \emptyset$$

puisqu'on ne pourra pas avoir de $\omega \in \Omega$ tel que $X(\omega) \in A$ et $X(\omega) \in B$. Comme par ailleurs

$$\{X \in A \cup B\} = \{X \in A\} \cup \{X \in B\} \ ,$$

on obtient que $P(X \in A \cup B) = P(X \in A) + P(X \in B)$ et donc, par définition de $P_X$, que $P_X(A \cup B) = P_X(A) + P_X(B)$. En conclusion, $P_X$ vérifie les deux propriétés voulues pour une probabilité. Le théorème est démontré. $\square$

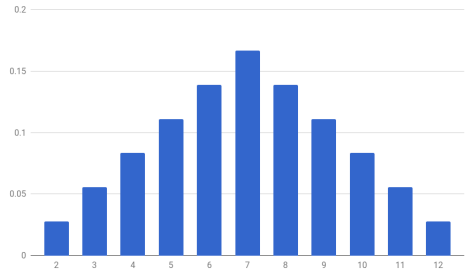
Pour caractériser $P_X$ il suffit de connaître les $P_X(\{x\})$ pour tous les $x \in E$. C'est essentiellement ce que nous dit le Théorème 13.1. On le répète ici, dans le contexte des variables aléatoires. On écrira aussi $P(X = x)$ pour $P_X(\{x\})$, la notation $X = x$ signifiant bien que l'on se réfère aux antécédants de $x$ par $X$. Donc, avec cette notation, $P(X = x) = P_X(\{x\})$.

**Théorème 22.2** (Caractérisation de $P_X$)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$. Soit $P_X$ la loi de $X$. Alors $P_X$ est entièrement caractérisée par les $P_X(\{x\}) = P(X = x)$ pour $x \in E$ via la relation*

$$P_X(A) = \sum_{x \in A} P(X = x)$$

*pour tout $A \in \mathcal{P}(E)$. Lorsque $E \subset \mathbb{Z}$ est un sous ensemble des relatifs $\mathbb{Z}$, on peut représenter $P_X$ par un histogramme.*

Un histogramme est un schéma en bâtons. Les abscisses sont les $x \in E$. Les ordonnées sont les $P_X(\{x\}) = P(X = x)$. Par exemple l'histogramme suivant représente la loi de la variable aléatoire "somme de deux dés".




HISTOGRAMME POUR LA SOMME DE DEUX DÉS

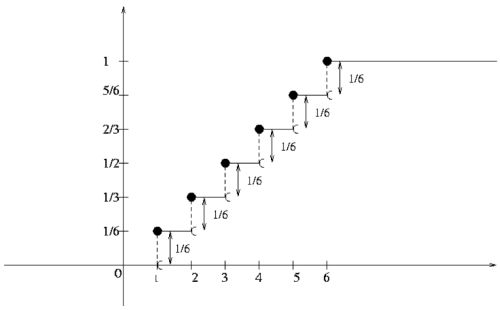
Dans la somme de deux dés, $\Omega = [\![1, 6]\!]^2$. Donc $\sharp\Omega = 36$. Si $X$ est la variable aléatoire somme des deux dés, alors $X : \Omega \to [\![2, 12]\!]$. Et on a $P(X = 2) = \frac{1}{36} \sim 0,028$, $P(X = 3) = \frac{2}{36} \sim 0,055$, $P(X = 4) = \frac{3}{36} \sim 0,083$, $P(X = 5) = \frac{4}{36} \sim 0,111$, $P(X = 6) = \frac{5}{36} \sim 0,138$, $P(X = 7) = \frac{6}{36} \sim 0,166$, $P(X = 8) = \frac{5}{36} \sim 0,138$, $P(X = 9) = \frac{4}{36} \sim 0,111$, $P(X = 10) = \frac{3}{36} \sim 0,083$, $P(X = 11) = \frac{2}{36} \sim 0,055$ et $P(X = 12) = \frac{1}{36} \sim 0,028$. On retrouve l'histogramme ci-dessus. Pour parler de fonction de répartition il faut arriver dans $\mathbb{R}$ pour avoir les relations d'ordre $\leq$, $<$, $\geq$ et $>$. On parle de variable aléatoire réelle.

**Définition 22.2** (Variables aléatoires réelles)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$. On dit que $X$ est une variable aléatoire réelle si $E \subset \mathbb{R}$.*

On prendra toujours $E$ fini ($\sharp E < +\infty$), dans notre contexte des variables aléatoires sur des espaces de probabilités finis. La fonction de répartition pour une variable aléatoire réelle est définie dans ce qui suit.

**Définition 22.3** (Fonction de répartition)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ avec $E \subset \mathbb{R}$. Soit $P_X$ la loi de $X$. On appelle fonction de répartition de $X$ la fonction réelle $F_X : \mathbb{R} \to [0, 1]$ définie par $F_X(x) = P_X(X \leq x)$, où l'on a écrit $P(X \leq x)$ pour $P(\{X \leq x\})$ et où $\{X \leq x\} = \{\omega \in \Omega \ / \ X(\omega) \leq x\}$.*

Soit $X$ la variable aléatoire correspondant au chiffre obtenu lors du lancé d'un dé. Alors $\Omega = [\![1,6]\!]$ et $P(X = x) = \frac{1}{6}$ pour $x \in [\![1,6]\!]$. La fonction de répartition $F_X : \mathbb{R} \to [0,1]$ est donnée par le graphe suivant




FONCTION DE RÉPARTITION POUR LE LANCÉ D'UN DÉ

On a $F_X(x) = 0$ si $x < 1$, $F_X(x) = \frac{1}{6}$ si $1 \leq x < 2$, $F_X(x) = \frac{1}{3}$ si $2 \leq x < 3$, $F_X(x) = \frac{1}{2}$ si $3 \leq x < 4$, $F_X(x) = \frac{2}{3}$ si $4 \leq x < 5$, $F_X(x) = \frac{5}{6}$ si $5 \leq x < 6$ et $F_X(x) = 1$ si $x \geq 6$.

**Théorème 22.3.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ avec $E \subset \mathbb{R}$. Soit $F_X : \mathbb{R} \to [0,1]$ la fonction de répartition de $X$. Alors $F_X$ est croissante, continue à droite et en escalier. De plus, pour tous réels $x < y$,*

$$P(x < X \leq y) = F_X(y) - F_X(x)$$

*et pour tout $x \in \mathbb{R}$, $P(X = x) = F_X(x) - F_X(x^-)$ où $F_X(x^-) = \lim_{y \to x, y < x} F_X(y)$. En particulier la connaissance de $F_X$ entraîne la connaissance de $P_X$ (et réciproquement bien sûr aussi puisque $F_X$ est définie à partir de $P_X$).*

*Démonstration.* Il est clair que $F_X$ est croissante en vertue de la propriété (2) du Théorème 12.1 puisque si $x \leq y$, $\{X \leq x\} \subset \{X \leq y\}$. Comme $\Omega$ est fini, $X$ ne prend qu'un nombre fini de valeurs. Si $x$ est l'une d'entres elles (à savoir s'il existe $\omega \in \Omega$ tel que $X(\omega) = x$) alors clairement, pour $\varepsilon > 0$ suffisamment petit, $[x, x+\varepsilon] \cap \mathrm{Im}(X) = \{x\}$, où $\mathrm{Im}(X) = \{X(\omega), \omega \in \Omega\}$. Par suite $\{X \leq y\} = \{X \leq x\}$ pour $y \in [x, x+\varepsilon]$ et donc $F_X(y) = F_X(x)$ pour $y \in [x, x+\varepsilon]$. En particulier $F_X$ est continue à droite en $x$ et $F_X$ est constante sur $[x, x+\varepsilon]$. Supposons maintenant que $x \notin \mathrm{Im}(X)$. Alors, clairement, pour $\varepsilon > 0$ suffisamment petit, $[x, x+\varepsilon] \cap \mathrm{Im}(X) = \emptyset$. La conclusion est la même: $F_X(y) = F_X(x)$ pour $y \in [x, x+\varepsilon]$, $F_X$ est continue à droite en $x$ et $F_X$ est constante sur $[x, x+\varepsilon]$. Si $x_1 < x_2 < \cdots < x_k$ sont les valeurs prises par $X$ on a en fait que $F_X(x) = 0$ si $x < x_1$, $F_X(x) = F_X(x_i)$ si $x_i \leq x < x_{i+1}$ et $F_X(x) = 1$ si $x \geq x_k$. De la même façon, on voit que si $x \in \mathrm{Im}(X)$, alors pour $\varepsilon > 0$, $[x-\varepsilon, x] \cap \mathrm{Im}(X) = \{x\}$. On en déduit que $F_X$ est constante sur $]x-\varepsilon, x[$ et comme $\{X \leq x\} = \{X = x\} \cup \{X < x\}$, l'union étant disjointe, on peut écrire que

$$F_X(x) = P(X = x) + F_X(y)$$

pour tout $y \in ]x-\varepsilon, x[$. On récupère bien $P(X = x) = F_X(x) - F_X(x^-)$. Supposons maintenant que $x \notin \mathrm{Im}(X)$, alors pour $\varepsilon > 0$, $[x-\varepsilon, x] \cap \mathrm{Im}(X) = \emptyset$. On en déduit que $F_X$ est constante sur $]x-\varepsilon, x]$ et comme $\{X \leq x\} = \{X = x\} \cup \{X < x\}$, l'union étant disjointe (intersection vide), on peut maintenant écrire tout à la fois

que $F_X(x) = P(X = x) + F_X(y)$ et que $F_X(x) = F_X(y)$ pour tout $y \in ]x - \varepsilon, x[$. On récupère alors que $P(X = x) = F_X(x) - F_X(x^-) = 0$. Considérons pour finir que $x < y$. Alors $\{X \leq y\} = \{x < X \leq y\} \cup \{X \leq x\}$ et l'union est disjointe. Donc $F_X(y) = P(x < X \leq y) + F_X(x)$. Le théorème est démontré. $\square$

## 23. Quelques exercices 5

**EXERCICE 20:** A l'oral du baccalauréat, les candidats tirent 3 sujets au sort. Il y a 100 sujets possibles proposés. Sur les 3 sujets tirés le candidat en choisit un. Pierre, qui passe le baccalauréat, a révisé 60 sujets sur les 100. Quelle est la probabilité pour que Pierre ait révisé les 3 sujets tirés ? Quelle est la probabilité pour que Pierre ait révisé exactement 2 sujets sur les trois tirés ? Quelle est la probabilité pour que Pierre ait révisé exactement 1 sujet sur les trois tirés ? Quelle est la probabilité que Pierre n'ait révisé aucun des trois sujets tirés ? Définir une variable aléatoire qui soit associée à ce problème. Quelle est sa loi ?

CORRECTION: Les sujets sont tirés au sort, nous sommes donc dans une situation d'équiprobabilité. L'univers des possibles $\Omega$ est composé des combinaisons de 3 sujets possibles. On a $\sharp\Omega = \binom{100}{3}$. Soit $A_i$ l'événement "Pierre a révisé exactement $i$ sujets sur les 3 tirés", $i = 0, 1, 2, 3$. Il y a exactement 40 sujets qui n'ont pas été révisés par Pierre. On a

$$\sharp A_3 = \binom{60}{3} \ , \ \sharp A_2 = \binom{2}{60}\binom{1}{40} \ ,$$
$$\sharp A_1 = \binom{1}{60}\binom{2}{40} \ , \ \sharp A_0 = \binom{3}{40} \ .$$

Ensuite $P(A_i) = \frac{\sharp A_i}{\sharp\Omega}$ et on trouve donc

$$P(A_0) = \frac{494}{8085} \sim 0,06 \ , \ P(A_1) = \frac{156}{539} \sim 0,29 \ ,$$
$$P(A_2) = \frac{236}{539} \sim 0,44 \ , \ P(A_3) = \frac{1711}{8085} \sim 0,21 \ .$$

Ce n'est pas demandé, mais la probabilité que Pierre ait révisé au moins un sujet est $1 - P(A_0) \sim 0,94$ et la probabilité que Pierre ait révisé au moins deux sujets est $P(A_2) + P(A_3) \sim 0,65$. On parle là de probabilité. Je ne suis pas certain qu'il soit conseillé de prendre le risque ... Quoiqu'il en soit, ces calculs répondent à la première série de questions. La variable aléatoire qui est associé au problème est la variable aléatoire $X : \Omega \to [\![0, 3]\!]$ qui à $\omega \in \Omega$ associe le nombre de sujets que Pierre a revisé dans $\omega$. La loi de $X$ est caractérisée par les $P_X(\{k\}) = P(X = k)$ pour $k \in [\![0, 3]\!]$ (cf. Théorème 22.2), et on a (cf. les calculs ci-dessus) que

$$P_X(k) = \frac{\binom{60}{k}\binom{40}{3-k}}{\binom{100}{3}} \ .$$

On le verra un peu plus loin, il s'agit d'une loi hypergéométrique de paramètres $(60, 40, 3)$ (cf. Section 26).

**EXERCICE 21:** On lance deux dés parfaitement équilibrés. On note $X$ le plus grand des numéros obtenus. Déterminer la loi de la variable aléatoire $X$.

CORRECTION: Les dés sont parfaitement équilibrés. Nous sommes dans une situation d'équiprobabilité. L'univers des possibles $\Omega$ est $\Omega = [\![1, 6]\!]^2$. On a $\sharp\Omega = 36$. La

variable aléatoire $X$ considérée est à valeurs dans $[\![1,6]\!]$. La loi de $X$ est caractérisée par les $P_X(\{k\}) = P(X = k)$ pour $k \in [\![1,6]\!]$ (cf. Théorème 22.2). On a

$$\{X = 1\} = \{(1,1)\} \ , \ \{X = 2\} = \{(1,2),(2,2),(2,1)\} \ ,$$
$$\{X = 3\} = \{(1,3),(2,3),(3,3),(3,2),(3,1)\} \ ,$$
$$\{X = 4\} = \{(1,4),(2,4),(3,4),(4,4),(4,3),(4,2),(4,1)\} \ ,$$
$$\{X = 5\} = \{(1,5),(2,5),(3,5),(4,5),(5,5),(5,4),(5,3),(5,2),(5,1)\} \ ,$$
$$\{X = 6\} = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(6,5),(6,4),(6,3),(6;2),(6,1)\} \ .$$

Donc $\sharp\{X = 1\} = 1$, $\sharp\{X = 2\} = 3$, $\sharp\{X = 3\} = 5$, $\sharp\{X = 4\} = 7$, $\sharp\{X = 5\} = 9$ et $\sharp\{X = 6\} = 11$. On a $P_X(i) = \frac{\sharp\{X=i\}}{36}$. On trouve alors

$$P_X(1) = \frac{1}{36} \ , \ P_X(2) = \frac{1}{12} \ , \ P_X(3) = \frac{5}{36} \ ,$$
$$P_X(4) = \frac{7}{36} \ , \ P_X(5) = \frac{1}{4} \ , \ P_X(6) = \frac{11}{36} \ .$$

Comme déjà dit, les $P_X(k)$ pour $k = 1, \ldots, 6$ caractérisent la loi de $X$.

**EXERCICE 22:** Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit de plus $X : \Omega \to [\![-2,2]\!]$ une variable aléatoire dont la loi est donnée par le tableau

| $k$ | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| $P_X(k)$ | 0,10 | 0,35 | 0,15 | 0,25 | 0,15 |

Vérifier que les $p_k = P_X(k)$ du tableau correspondent bien à une probabilité sur $[\![-2,2]\!]$. Déterminer la loi de $X^2$.

CORRECTION: Soit $p_k = P_X(k)$. On vérifie que l'on a bien une probabilités au sens du Théorème 13.1, et donc que l'on a bien que $\sum_{k=-2}^{2} p_k = 1$. C'est effectivement le cas: $0,10 + 0,35 + 0,15 + 0,25 + 0,15 = 1$. On remarque maintenant que $X^2 : \Omega \to \{0,1,4\}$. Il nous faut calculer $P(X^2 = 0)$, $P(X^2 = 1)$ et $P(X^2 = 4)$. On a $\{X^2 = 0\} = \{X = 0\}$, $\{X^2 = 1\} = \{X = -1\} \cup \{X = 1\}$ et l'union est disjointe, et $\{X^2 = 4\} = \{X = -2\} \cup \{X = 2\}$ et, là encore, l'union est disjointe. On a alors $P(X^2 = 0) = P(X = 0) = 0,15$ puis, par additivité, on peut écrire que $P(X^2 = 1) = P(X = -1) + P(X = 1) = 0,60$ et enfin, toujours par additivité, $P(X^2 = 4) = P(X = -2) + P(X = 2) = 0,25$. La loi de $X^2$ est caractérisée par les $P(X^2 = k)$ pour $k = 0,1,4$ (cf. Théorème 22.2).

## 24. Deux lois discrètes finies usuelles

On entame avec cette section le catalogue des lois discrètes que l'on rencontre le plus souvent. On discutera dans cette section des lois uniformes et de Bernoulli.

**Définition 24.1** (Loi uniforme)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ avec $E = \{x_1, \ldots, x_n\}$. On dit que $X$ suit une loi discrète uniforme si $P(X = x_i) = \frac{1}{n}$ pour tout $i = 1, \ldots, n$. Lorsque $x_i = i$ pour tout $i = 1, \ldots, n$ (donc en particulier $E \subset \mathbb{N}$) on dit que $X$ suit la loi uniforme discrète de paramètre $n$. On la note $\mathcal{U}(n)$.*

L'exemple pratique type d'une variable aléatoire qui suit une loi uniforme discrète est donné par le lancer de dé. La variable aléatoire $X$ correspondant au chiffre obtenu suit la loi uniforme discrète $\mathcal{U}(6)$ de paramètre 6.

**Définition 24.2** (Loi de Bernoulli)**.** *Soit $p \in [0,1]$ un réel. Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to \{0,1\}$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ qui ne prend que deux valeurs 0 et 1. On dit que $X$ suit la loi de Bernoulli de paramètre $p$ si $P(X=1) = p$ et $P(X=0) = 1-p$. On note $\mathcal{B}(p)$ la loi de Bernoulli de paramètre $p$.*

Un exemple typique d'une variable aléatoire qui suit une loi de Bernoulli correspond au lancé d'une pièce. Si celle-ci est bien équilibrée, $p = \frac{1}{2}$, sinon on peut avoir $p \neq \frac{1}{2}$. De manière générale, la loi de Bernoulli est la loi d'une variable aléatoire qui code le résultat d'une épreuve qui n'admet que deux issues: 1 pour "succès", 0 pour "échec", ou quel que soit le nom qu'on donne aux deux issues d'une telle expérience aléatoire. La preuve de cette affirmation est l'objet du théorème suivant.

**Théorème 24.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$. Si $\sharp E = 2$ alors $X$ suit une loi de Bernoulli $\mathcal{B}(p)$ pour un certain $p \in [0,1]$.*

*Démonstration.* $E$ contient deux éléments. Appelons-les 0 et 1. Clairement donc $\Omega = \{X = 0\} \cup \{X = 1\}$ et cette union est disjointe (l'intersection des deux ensembles est vide). Donc $1 = P(\Omega) = P(X=0) + P(X=1)$. Notons maintenant $p = P(X=1)$. Alors $P(X=0) = 1-p$ et $X$ suit la loi $\mathcal{B}(p)$. Le théorème est démontré. □

Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A \in \mathcal{P}(\Omega)$ un événement quelconque. Notons $\mathbb{1}_A$ la fonction indicatrice de $A$, donc la fonction $\mathbb{1}_A : \Omega \to \mathbb{R}$ définie par $\mathbb{1}_A(\omega) = 0$ si $\omega \notin A$ et $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$. Alors $\mathbb{1}_A$ est une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ qui ne prend que deux valeurs 0 et 1. Elle suit ainsi une loi de Bernoulli $\mathcal{B}(p)$ de paramètre $p = P(A)$.

## 25. La reine des lois: la loi binomiale

La loi binomiale correspond à l'expérience consistant à lancer $n$ fois une pièce. La pièce n'étant pas forcément parfaitement équilibrée on va avoir une probabilité $p$ de tomber sur pile, $p \in [0,1]$, et du coup une probabilité $1-p$ de tomber sur face. Avec $n$ lancers l'univers des possibles $\Omega$ est donné par $\Omega = \{P, F\}^n$. Les lancers de la pièce sont indépendants les uns des autres. Les événements $A_1, \ldots, A_n$ où $A_i$ est l'événement "obtenir $F$ au ième lancer" (ou obtenir $P$, cela revient au même) sont donc mutuellement indépendants (on peut se référer à la formule des probabilités composées avec comme information que $P(A_2/A_1) = P(A_2)$ puisque le second tirage se "fiche" du premier, que $P(A_3/A_1 \cap A_2) = P(A_3)$ puisque le troisième se "fiche" des deux premiers etc.) Si $\omega \in \Omega$ comporte $k$ côtés piles, et donc $n-k$ côtés faces, on a alors

$$P(\{\omega\}) = p^k (1-p)^{n-k} \ .$$

Soit $X : \Omega \to [\![0, n]\!]$ la variable aléatoire qui à $\omega \in \Omega$ fait correspondre

$$X(\omega) \ = \ \text{le nombre de côtés piles que contient } \omega \ .$$

Soit aussi $A_k \in \mathcal{P}(\Omega)$ l'événement constitué des $\omega$ qui comportent exactement $k$ côtés piles. Pour dénombrer $A_k$ on veut savoir où vont se situer les $k$ piles. Les emplacements étant numérotés de 1 à $n$ ils sont donnés par les sous ensembles à $k$

éléments de $[\![1,n]\!]$. On parle donc de combinaisons. Donc

$$\sharp A_k = \binom{n}{k} .$$

Pour tous $\omega, \omega' \in A_k$, $P(\{\omega\}) = P(\{\omega'\})$. Donc (cf. Théorème 13.1)

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

pour tout $k \in [\![0,n]\!]$. On a là la loi du nombre de succès (du nombre de piles dans notre cas) dans un jeu de pile ou face où l'on lance $n$ fois de suite une pièce pas forcément bien équilibrée (ou pas..., $p = 1/2$ si la pièce est parfaitement équilibrée).

**Définition 25.1** (Loi binomiale)**.** *Soient $n \in \mathbb{N}^\star$ et $p \in [0,1]$. Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to [\![0,n]\!]$ une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ qui prend les valeurs $0, 1, \ldots, n$. On dit que $X$ suit la loi binomiale de paramètres $n$ et $p$ si*

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

*pour tout $k \in [\![0,n]\!]$. On note $\mathcal{B}(n,p)$ la loi binomiale de paramètres $n$ et $p$. La formule du binôme (cf. Théorème 7.1) assure que la somme des probabilités élémentaires vaut 1 puisque $p + (1-p) = 1$.*

On pourra remarquer que $\mathcal{B}(1,p) = \mathcal{B}(p)$. On retrouve la loi de Bernoulli lorsque $n = 1$. Comme pour la loi de Bernoulli, on rencontre la loi binomiale dans de nombreuses configurations.

**Théorème 25.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $A_1, \ldots, A_n$ une famille indépendante d'événements. On suppose que les $P(A_i)$ sont tous égaux. On note $p = P(A_1) = \cdots = P(A_n)$. Soit $X : \Omega \to [\![0,n]\!]$ la variable aléatoire*

$$X = \sum_{i=1}^{n} \mathbb{1}_{A_i}$$

*qui compte le nombre d'événements $A_i$ réalisés. Alors $X$ suit la loi binomiale $\mathcal{B}(n,p)$ de paramètres $n$ et $p$.*

*Démonstration.* On peut suivre la même analyse que dans la discussion du début de la section. Soit $k \in [\![0,n]\!]$ donné. L'événement $\{X = k\}$ est donné par la réunion de toutes les intersections possibles de $A_i$ et $A_i^c$ comportant exactement $k$ événements $A_i$. Notons $\Theta = \{0,1\}$. Pour $j = 1, \ldots, n$ et $i_j \in \Theta$ on pose $B_{i_j} = A_j$ si $i_j = 0$ et $B_{i_j} = A_j^c$ si $i_j = 1$. Alors

$$\{X = k\} = \bigcup_{(i_1,\ldots,i_n) \in \Theta_k^n} B_{i_1} \cap \cdots \cap B_{i_n} ,$$

où $\Theta_k^n$ est l'ensemble des $(i_1, \ldots, i_n)$ dans $\Theta^n$ tels que $k$ exactement des $i_j$ sont des 0. La réunion est disjointe car pour deux $n$-uplets distincts $(i_1, \ldots, i_n)$ et $(i'_1, \ldots, i'_n)$ dans $\Theta^n$, les $B_{i_1} \cap \cdots \cap B_{i_n}$ et $B_{i'_1} \cap \cdots \cap B_{i'_n}$ seront d'intersections vides (si les deux $n$-uplets sont distincts il va y avoir un $i_j \neq i'_j$ pour un $j$ et l'on va se retrouver dans l'intersection des deux intersections avec un $A_j \cap A_j^c = \emptyset$). Comme les $A_1, \ldots, A_n$ sont indépendants (mutuellement indépendants) on obtient avec la Définition 20.2 et le Théorème 20.2 que pour tout $(i_1, \ldots, i_n) \in \Theta_k^n$,

$$P(B_{i_1} \cap \cdots \cap B_{i_n}) = p^k (1-p)^{n-k}$$

et on aura ensuite que

$$P(\{X = k\}) = (\sharp\Theta_k^n)\, p^k (1-p)^{n-k} \ .$$

Pour calculer le cardinal $\sharp\Theta_k^n$ de $\Theta_k^n$ on procède comme en début de section. On cherche le nombre d'emplacements possibles des 0. Les emplacements étant numérotés de 1 à $n$ ils sont donnés par les sous ensembles à $k$ éléments de $[\![1, n]\!]$. On parle donc de combinaisons. Donc $\sharp\Theta_k^n = \binom{n}{k}$. Le théorème est démontré. $\square$

## 26. La loi hypergéométrique

Le cas de figure type ici est celui d'une urne où se trouvent $m_1$ boules noires et $m_2$ boules blanches. Pour des raisons pratiques de dénombrement, les boules sont considérées toutes différentes (distinguables), même si certaines sont identiques. Grâce à cette hypothèse, la probabilité de tirer une boule est la même pour toutes les boules. Pour $n \leq \min(m_1, m_2)$, on tire sans remise $n$ boules de l'urne. On note $X$ le nombre de boules noires tirées. Si $\Omega$ est l'univers des possibles, alors $\Omega$ est constitué des combinaisons possibles de $n$ éléments dans un esnemble de $m = m_1 + m_2$ éléments. Donc $\sharp\Omega = \binom{m}{n}$. Et $X$ est la variable aléatoire $X : \Omega \to [\![0, n]\!]$ qui à $\omega \in \Omega$ associe le nombre de boules noires que contient $\omega$. Soit $k \in [\![0, n]\!]$. Clairement $\{X = k\}$ est constitué des combinaisons possibles de $k$ éléments dans $m_1$ et de $n - k$ éléments dans $m_2$. Donc $\sharp\{X = k\} = \binom{m_1}{k}\binom{m_2}{n-k}$. Au final,

$$P(X = k) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}} \tag{26.1}$$

pour tout $k \in [\![0, n]\!]$. On parle de loi hypergéométrique de paramètres $(m_1, m_2, n)$. Etant donné $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini, on dit qu'une variable aléatoire $X$ suit une loi hypergéométrique de paramètres $(m_1, m_2, n)$ si sa loi est donnée par (26.1).

## 27. Modélisation et lois usuelles

On discute quelques applications courantes et les lois auxquelles elles se réfèrent. N'hésitez pas à agrandir la liste avec ce que vous pouvez rencontrer dans vos exercices.

EXPÉRIENCE 1: *"On lance un dé non pipé et on relève le chiffre obtenu"*. L'univers des possibles $\Omega = [\![1, 6]\!]$ et $X : \Omega \to [\![1, 6]\!]$ est l'identité. La variable aléatoire $X$ suit la loi discrète uniforme de paramètre 6.

EXPÉRIENCE 2: *"Une urne contient 12 boules noires, 15 boules blanches, et 9 boules bleues. On effectue 3 tirages successifs avec remise. On compte le nombre de boules bleues obtenues"*. On est là dans une succession d'épreuves de Bernoulli décrite par le Théorème 25.1. Soit $\mathbb{B}$ l'ensemble des 36 boules. L'univers des possibles $\Omega$ est $\Omega = \mathbb{B}^3$. On a $\sharp\Omega = (\sharp\mathbb{B})^3$. Soit $A_i$ l'événement "obtenir une boule bleue au $i$ème tirage". Si $\mathbb{B}_\ell \subset \mathbb{B}$ est l'ensemble des boules bleues, alors $A_1 = \mathbb{B}_\ell \times \mathbb{B} \times \mathbb{B}$, $A_2 = \mathbb{B} \times \mathbb{B}_\ell \times \mathbb{B}$ et $A_3 = \mathbb{B} \times \mathbb{B} \times \mathbb{B}_\ell$. On est dans une situation d'équiprobabilité et les événements $A_i$ sont clairement indépendants. La variable aléatoire $X$ qui compte le nombre de boules bleues après 3 tirages est $X = \sum_{i=1}^3 \mathbb{1}_{A_i}$. On a $P(A_i) = (\sharp\mathbb{B})^2\sharp\mathbb{B}_\ell/\sharp\Omega = \sharp\mathbb{B}_\ell/\sharp\mathbb{B}$ et donc $P(A_i) = 1/4$ pour tout $i$. D'après le Théorème 25.1, $X$ suit la loi binomiale de paramètres 3 et $\frac{1}{4}$.

EXPÉRIENCE 3: *"On range au hasard 10 jetons dans 3 sacs. On compte le nombre de jetons dans le premier sac"*. C'est le même principe que pour l'expérience 2. On place un premier jeton dans l'un des trois cas au hasard, puis un second, puis un troisième etc. Comme à l'expérience 2 on a là une suite d'expériences de Bernoulli, avec à chaque fois une probabilité 1/3 d'être dans le premier sac. On récupère ainsi la loi binomiale de paramètres 10 et $\frac{1}{3}$ pour la variable aléatoire qui compte le nombre de jetons dans le premier sac.

EXPÉRIENCE 4: *"On range les 54 cartes d'un jeu de cartes en ligne, faces cachées. On retourne les cartes une à une jusqu'à arriver au valet de cœur. On compte le nombre de cartes retournées"*. La formulation ici est trompeuse. La variable aléatoire "nombre de cartes retournées pour arriver au valet de cœur" a pour loi élémentaire $P(X = k)$, pour $k = 1, \ldots, 54$, qui est la probabilité d'obtenir le valet de cœur à la $k$ème carte retournée, et donc la probabilité que le valet de cœur se situe à la $k$ème position. Les cartes sont alignées au hasard. La probabilité que le valet de cœur soit à la première place, à la 5ème ou à la 32ème sont les mêmes. Donc $X$ suit la loi discrète uniforme de paramètre 54.

EXPÉRIENCE 5: *"Un QCM comporte 20 questions, et pour chaque question il y a 5 choix possibles. Un candidat répond au hasard. On compte le nombre de bonnes réponses"*. Il s'agit d'une suite de 20 expériences de Bernoulli, et pour chacune d'entre elle la probabilité d'avoir une réponse juste est 1/5. La variable aléatoire $X$ qui compte le nombre de bonnes réponses suit une loi binomiale de paramètres 20 et $\frac{1}{5}$.

EXPÉRIENCE 6: *"Un joueur au loto coche 6 numéros parmi 49. Les 6 numéros gagnants sont déterminés par tirage au sort. On s'intéresse au nombre de numéros gagnants sur le bulletin"*. L'univers des possibles $\Omega$ est ici l'ensemble des familles non ordonnées de 6 élément parmi 49. On parle de combinaisons et $\sharp\Omega = \binom{49}{6}$. La variable aléatoire $X$ correspondant au nombre de numéros gagnants est à valeurs dans $[\![0, 6]\!]$. Soit $k \in [\![0, 6]\!]$. Alors $P(X = k)$ se calcule par équiprobabilité. Le nombre de cas favorables étant donnés par le choix de $k$ boules parmi 6 et de $6 - k$ boules parmi 43. On a donc

$$P(X = k) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$$

pour tout $k \in [\![0, 6]\!]$. Donc $X$ suit la loi hypergéométrique de paramètres $(6, 43, 6)$. Cette version du loto a été utilisée de 1976 à 2008.

EXPÉRIENCE 7: *"A l'oral du baccalauréat, les candidats tirent* 3 *sujets au sort. Il y a 100 sujets possibles proposés. Sur les* 3 *sujets tirés le candidat en choisit un. Un candidat qui passe le baccalauréat a révisé 60 sujets sur les 100. On s'intéresse au nombre de sujets révisés par le candidat parmi les sujets tirés"*. La loi de la variables aléatoire correspondante est la loi hypergéométrique de paramètres $(60, 40, 3)$ (cf. Exercice 20).

## 28. Espérance

On considère ici des variables aléatoires réelles, à savoir à valeurs dans une sous ensemble fini $E$ de $\mathbb{R}$. On peut attacher à ces variables aléatoires plusieurs quantités, à commencer par la moyenne pondérée des valeurs que peut prendre cette variable, la moyenne étant pondérée par les probabilités d'apparition de chaque

valeur. On parle d'espérance en référence à la notion d'espérance de vie utilisée dans le langage des assurances. L'espérance mathématique d'une variable aléatoire réelle est, intuitivement, la valeur que l'on s'attend à trouver, en moyenne, si l'on répète un grand nombre de fois la même expérience aléatoire (c'est ce que nous dira la loi des grands nombres). On trouve souvent l'interprétation qu'un jeu est équitable lorsque l'espérance est nulle, en faveur du joueur lorsque l'espérance est positive et en défaveur du joueur lorsque l'espérance est négative.

**Définition 28.1** (Espérance)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire réelle. L'espérance de $X$, notée $\mathbb{E}(X)$, est définie par*

$$\begin{aligned} \mathbb{E}(X) &= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \\ &= \sum_{x \in E} x P_X(x) \ , \end{aligned} \tag{28.1}$$

*où, par abus de langage, on note $P_X(x)$ pour $P_X(\{x\})$ et où $P_X$ est la loi de $X$ de sorte que $P_X(x) = P(X = x)$. En particulier, l'espérance de $X$ ne dépend que de sa loi. Une variable aléatoire $X$ est dite centrée lorsque $\mathbb{E}(X) = 0$.*

Il y a quand même une affirmation qu'il faut démontrer dans cette proposition, à savoir l'égalité entre les deux lignes de (28.1). Elle découle du Théorème 13.1. On a clairement $\Omega = \cup_{x \in E} \{X = x\}$ et l'union est disjointe (les $\{X = x\}$ pour $x \in E$ sont deux à deux disjoints). En vertue du Théorème 13.1,

$$P(X = x) = \sum_{\omega \in \{X = x\}} P(\{\omega\})$$

tandis que pour $\omega \in \{X = x\}$, $X(\omega) = x$. On en déduit que

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) &= \sum_{x \in E} \left( \sum_{\omega \in \{X = x\}} X(\omega) P(\{\omega\}) \right) \\ &= \sum_{x \in E} x \sum_{\omega \in \{X = x\}} P(\{\omega\}) \\ &= \sum_{x \in E} x P_X(x) \ , \end{aligned}$$

ce qui est l'égalité voulue et annoncée dans la formule (28.1). Des propriétés simples de l'expérance sont données dans les trois propositions suivantes. La linéarité de l'espérance est démontrée dans la proposition qui suit.

**Proposition 28.1** (Linéarité de l'espérance)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soient $X : \Omega \to E$ et $Y : \Omega \to E'$ deux variables aléatoires réelles. Soient $\lambda, \mu \in \mathbb{R}$. Alors $Z = \lambda X + \mu Y$ est encore une variable aléatoire réelle sur $(\Omega, \mathcal{P}(\Omega), P)$ et $\mathbb{E}(Z) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y)$.*

*Démonstration.* Par définition de l'espérance,

$$\begin{aligned}\mathbb{E}(Z) &= \sum_{\omega\in\Omega} Z(\omega)P(\{\omega\}) \\ &= \lambda \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}) + \mu \sum_{\omega\in\Omega} Y(\omega)P(\{\omega\}) \\ &= \lambda\mathbb{E}(X) + \mu\mathbb{E}(Y)\end{aligned}$$

puisque $Z(\omega) = \lambda X(\omega) + \mu Y(\omega)$, et le théorème est démontré. □

Pour $a$ une constante, $\mathbb{E}(a) = a$. Par linéarité de l'espérance, pour toute variable aléatoire $X$, $X - \mathbb{E}(X)$ est une variable aléatoire centrée. La positivité de l'espérance est démontrée dans la proposition qui suit.

**Proposition 28.2** (Positivité de l'espérance)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soient $X : \Omega \to E$ et $Y : \Omega \to E'$ deux variables aléatoires réelles. Si $Y \geq X$ (au sens des fonctions, donc $Y(\omega) \geq X(\omega)$ pour tout $\omega \in \Omega$) alors $\mathbb{E}(Y) \geq \mathbb{E}(X)$.*

*Démonstration.* Si $Y \geq X$ alors $Y - X \geq 0$ et donc $\mathbb{E}(Y - X) \geq 0$ puisque $(Y - X)(\omega) \geq 0$ pour tout $\omega \in \Omega$. Par linéarité de l'espérance (Proposition 28.1) on récupère que $\mathbb{E}(Y) - \mathbb{E}(X) \geq 0$ et donc que $\mathbb{E}(Y) \geq \mathbb{E}(X)$. Le théorème est démontré. □

La troisième et dernière des propriétés démontrées dans cette section est dite de transport. Si $X : \Omega \to E$ est une variable aléatoire (non nécessairement réelle), et $f : E \to \mathbb{R}$ est une fonction, on dit que $f(X)$ est la variable $X$ transportée par $f$.

**Proposition 28.3** (Transport de l'espérance)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire (non forcément réelle). Soit $f : E \to \mathbb{R}$ une fonction. Alors $Y = f(X)$ est une variable aléatoire réelle sur $(\Omega, \mathcal{P}(\Omega), P)$ et $\mathbb{E}(Y) = \sum_{x\in E} f(x)P_X(x)$.*

*Démonstration.* On a $Y : \Omega \to f(E)$ où $f(E) = \mathrm{Im}(f) = \{f(x), x \in E\}$. Pour $y \in f(E)$ on a $\{Y = y\} = \bigcup_{\{x/f(x)=y\}} \{X = x\}$ et la réunion est disjointe. Donc $P(Y = y) = \sum_{\{x/f(x)=y\}} P(X = x)$. Avec (28.1) on pourra ainsi écrire que

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{y\in f(E)} yP(Y = y) \\ &= \sum_{y\in f(E)} y \sum_{x/f(x)=y} P(X = x) \\ &= \sum_{y\in f(E)} \sum_{x/f(x)=y} f(x)P(X = x) \\ &= \sum_{x\in E} f(x)P_X(x)\end{aligned}$$

car, comme on pourra s'en convaincre facilement, $E = \bigcup_{y\in f(E)} \{x \in E \ / \ f(x) = y\}$. La proposition est démontrée. □

## 29. Variance et écart type

La variance d'une variable aléatoire réelle est une mesure de la dispersion de ses valeurs. Elle mesure l'écart quadratique moyen d'une variable aléatoire réelle à son espérance. Si $X : \Omega \to E$ est cette variable aléatoire, on a, en vertue de la Proposition 28.3 avec $f(x) = x^2$,

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] \\ &= \sum_{x \in E} (x - \mathbb{E}(X))^2 \, P(X = x) \ . \end{aligned} \tag{29.1}$$

En d'autres termes la variance mesure de combien les valeurs prises par la variable aléatoire s'écartent de la valeur moyenne $\mathbb{E}(X)$. L'écart type est la racine carré de la variance. Par "homogénéité" (la variance renvoie au carré de la variable aléatoire) c'est plutôt lui qui donne l'ordre de grandeur des fluctuations autour de la moyenne.

**Définition 29.1** (Variance)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire réelle. La variance de $X$, notée $Var(X)$, est le nombre positif*

$$Var(X) = \mathbb{E}\big[(X - \mathbb{E}(X))^2\big]$$

*dont la formule calculatoire est donnée par* (29.1)*. L'écart type de $X$ est le nombre positif $\sigma(X) = \sqrt{Var(X)}$. Une variable aléatoire est dite réduite si son écart type (donc sa variance) vaut* 1*.*

Les variances vérifient un certain nombre de propriétés simples données par les propositions suivantes.

**Proposition 29.1** (Formule de Huygens)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire réelle. Si $Var(X)$ est la variance de $X$, on a aussi $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.*

*Démonstration.* On part de la formule (29.1). On a

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{x \in E} (x - \mathbb{E}(X))^2 \, P(X = x) \\ &= \sum_{x \in E} x^2 P(X = x) - 2\mathbb{E}(X) \sum_{x \in E} x P(X = x) + \mathbb{E}(X)^2 \sum_{x \in E} P(X = x) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 P(\Omega) \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

puisque $P(\Omega) = 1$. La proposition est démontrée. □

La proposition suivante suit de la linéarité de l'espérance. Elle donne la variance d'une variable aléatoire transportée par une fonction affine (donc $Y = f(X)$ avec $f(x) = ax + b$ une fonction affine).

**Proposition 29.2.** *Soient $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et $X : \Omega \to E$ une variable aléatoire réelle. Soient $a, b \in \mathbb{R}$. On a $Var(aX + b) = a^2 \, Var(X)$.*

*Démonstration.* On utilise la linéarité de l'espérance donnée à la Proposition 28.1 et, par exemple, la formule de Huygens de la Proposition 29.1. On utilise aussi que

pour une variable aléatoire constante $b$ (qui ne prend qu'une seule valeur, la valeur $b$), $\mathbb{E}(b) = bP(b = b) = bP(\Omega) = b$. On a alors

$$\begin{aligned}
\mathrm{Var}(aX + b) &= \mathbb{E}[(aX + b)^2] - \mathbb{E}(aX + b)^2 \\
&= \mathbb{E}(a^2X^2 + 2abX + b^2) - (a\mathbb{E}(X) + b)^2 \\
&= a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2 - a^2\mathbb{E}(X)^2 - 2ab\mathbb{E}(X) - b^2 \\
&= a^2\mathbb{E}(X^2) - a^2\mathbb{E}(X)^2 \\
&= a^2\mathrm{Var}(X)
\end{aligned}$$

et la proposition est démontrée. □

La troisième des propriétés revient sur l'interprétation de la variance dans le cas extrême $\mathrm{Var}(X) = 0$. Dans ce cas extrême, il n'y a pas de dispersion.

**Proposition 29.3.** *Soient $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et $X : \Omega \to E$ une variable aléatoire réelle. On suppose que la variance de $X$ est nulle, donc que $Var(X) = 0$. Alors $P(X = \mathbb{E}(X)) = 1$ ou, de façon équivalente, $P(S) = 0$ pour $S = \{\omega \ / \ X(\omega) \neq \mathbb{E}(X)\}$.*

*Démonstration.* En utilisant la formule (29.1), et comme $\mathrm{Var}(X) = 0$,

$$\sum_{x \in E} (x - \mathbb{E}(X))^2 \, P(X = x) = 0 \ .$$

Tous les termes de la somme sont positifs ou nuls et donc, si $x \neq \mathbb{E}(X)$, alors $P(X = x) = 0$. On a

$$S = \bigcup_{\{x \neq \mathbb{E}(X)\}} \{X = x\}$$

et la réunion est disjointe. Donc

$$P(S) = \sum_{\{x \neq \mathbb{E}(X)\}} P(X = x) = 0 \ .$$

On a

$$\Omega = \{X = \mathbb{E}(X)\} \cup S$$

et la réunion est disjointe. Donc

$$1 = P(X = \mathbb{E}(X)) + P(S) = P(X = \mathbb{E}(X)) \ .$$

La proposition est démontrée. □

## 30. Quelques exercices 6

**EXERCICE 23:** Soit $X$ une variable aléatoire à valeurs dans $\{0, 1, 2\}$. On suppose que $\mathbb{E}(X) = 1$ et $\mathrm{Var}(X) = \frac{1}{2}$. Déterminer la loi de $X$.

CORRECTION: Il nous faut calculer les probabilités élémentaires $p_0 = P(X = 0)$, $p_1 = P(X = 1)$ et $p_2 = P(X = 2)$. On sait que $p_0 + p_1 + p_2 = 1$ puisqu'il s'agit d'une probabilité, que $p_1 + 2p_2 = 1$ puisque $\mathbb{E}(X) = 1$ et que $p_1 + 4p_2 = \frac{3}{2}$ puisque $\mathbb{E}(X) = 1$ et $\mathrm{Var}(X) = \frac{1}{2}$ de sorte que $\mathbb{E}(X^2) = \frac{3}{2}$. On a donc le système

$$\begin{cases} p_0 + p_1 + p_2 = 1 \\ p_1 + 2p_2 = 1 \\ p_1 + 4p_2 = \frac{3}{2} \end{cases}$$

La 3ème équation à laquelle on soustrait la 2nde donne $2p_2 = \frac{1}{2}$ et donc $p_2 = \frac{1}{4}$. Avec la seconde équation on trouve ensuite $p_1 = \frac{1}{2}$. La première équation donne enfin $p_0 = \frac{1}{4}$. Ces quantités sont bien toutes positives ou nulles et, pour être tout à fait formel, on pourra vérifier que $(p_0, p_1, p_2) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ répond bien à question (à savoir vérifie les trois équations imposées).

**EXERCICE 24⋆:** On dispose de $n$ urnes numérotées de 1 à $n$. La première urne contient une boule notée 1, la seconde urne contient deux boules notées 1 et 2,..., la $n$ème urne contient $n$ boules numérotées de 1 à $n$. On choisit d'abord une urne puis ensuite une boule dans cette urne. On note $X$ la variable aléatoire du numéro obtenu. Quelle est la loi de $X$ ? Que vaut son espérance ?

CORRECTION: L'univers des possibles est l'ensemble des couples $(urne, boule)$, donc $[\![1, n]\!]^2$. On a $X : \Omega \to [\![1, n]\!]$. Notons $U$ la variable aléatoire du numéro de l'urne obtenue. On veut calculer les probabilités élémentaires $P(X = k)$. Les évéments $\{U = i\}$ pour $i = 1, \ldots, n$ forment un système complet d'évéments. On a bien sur que $P(U = i) = \frac{1}{n}$ puisque le choix de l'urne se fait au hasard. La formule généralisée des probabilités totales de la Proposition 16.2 donne que

$$\begin{aligned} P(X = k) &= \sum_{i=1}^{n} P(\{X = k\}/\{U = i\})P(U = i) \\ &= \frac{1}{n} \sum_{i=1}^{n} P(\{X = k\}/\{U = i\}) \ . \end{aligned}$$

On a $P(\{X = k\}/\{U = i\}) = 0$ si $k > i$ et $P(\{X = k\}/\{U = i\}) = \frac{1}{i}$ si $k \leq i$. Donc

$$P(X = k) = \sum_{i=k}^{n} \frac{1}{ni} \ .$$

Les $P(X = k)$ caractérisent la loi de $X$. Pour calculer l'espérance on applique la définition et on permute les sommes (ce qui est toujours possible puisqu'il s'agit de sommes finies). On note $\delta_{ik} = 1$ si $i \geq k$ et $\delta_{ik} = 0$ si $i < k$. On écrit tout d'abord que

$$\mathbb{E}(X) = \sum_{k=1}^{n} kP(X = k) = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=k}^{n} \frac{k}{i} = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{k}{i} \delta_{ik} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{k}{i} \delta_{ik} \ .$$

On remarque qu'avec la formule (31.1) de la somme des $n$ premiers entiers

$$\sum_{k=1}^{n} \frac{k}{i} \delta_{ik} = \frac{1}{i} \sum_{k=1}^{i} k = \frac{i+1}{2} \ .$$

Par suite, toujours avec (31.1),

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{2n} \sum_{i=1}^{n} (i+1) \\ &= \frac{1}{2n} \left( \frac{n(n+1)}{2} + n \right) \end{aligned}$$

et on trouve que $\mathbb{E}(X) = \frac{n+3}{4}$.

**EXERCICE 25⋆:** Soit $X$ une variable aléatoire à valeurs dans $[\![0, n]\!]$. Montrer que $\mathbb{E}(X) = \sum_{k=0}^{n-1} P(X > k)$.

CORRECTION: L'astuce consiste à écrire que $P(X=k)=P(X>k-1)-P(X>k)$ qui suit de l'union disjointe $\{X>k-1\}=\{X>k\}\cup\{X=k\}$. On obtient alors un télescopage dans les sommes:

$$\begin{aligned}\mathbb{E}(X)&=\sum_{k=1}^{n}kP(X>k-1)-\sum_{k=1}^{n}kP(X>k)\\&=\sum_{k=0}^{n-1}(k+1)P(X>k)-\sum_{k=1}^{n}kP(X>k)\\&=\sum_{k=0}^{n-1}P(X>k)\end{aligned}$$

car $P(X>n)=0$ dans la seconde somme.

**EXERCICE 26:** On considère un dé truqué à 6 faces. On note $X$ la variable aléatoire du numéro obtenu. On suppose que le dé est truqué de sorte que la probabilité d'obtenir une face est proportionnelle au numéro inscrit sur cette face. Déterminer la loi de $X$. Calculer son espérance et sa variance. On pose $Y=\frac{1}{X}$. Déterminer la loi de $Y$, son espérance et sa variance.

CORRECTION: La variable aléatoire $X$ est à valeurs dans $[\![1,6]\!]$. Par hypothèse il existe $\theta>0$ tel que $P(X=k)=\theta k$ pour tout $k\in[\![1,6]\!]$. On doit avoir $\sum_{k=1}^{6}P(X=k)=1$ et donc, comme $1+2+\cdots+6=21$, on trouve que $\theta=\frac{1}{21}$. On a alors $P(X=k)=\frac{k}{21}$. On calcule $\mathbb{E}(X)=\frac{1}{21}\sum_{k=1}^{6}k^2=\frac{13}{3}$ (on pourra utiliser la formule (31.2) ou alors calculer à la main). Pour la variance il faut calculer $\mathbb{E}(X^2)$. La variable aléatoire $X^2$ prend les valeurs $\{1,4,9,16,25,36\}$ et donc

$$\begin{aligned}\mathbb{E}(X^2)&=P(X=1)+4P(X^2=4)+9P(X^2=9)\\&\quad+16P(X^2=16)+25P(X^2=25)+36P(X^2=36)\\&=P(X=1)+4P(X=2)+9P(X=3)\\&\quad+16P(X=4)+25P(X=5)+36P(X=6)\\&=21\ .\end{aligned}$$

Par suite $\mathrm{Var}(X)=21-(\frac{13}{3})^2=\frac{20}{9}$. On s'intéresse maintenant à la loi de $Y$. La variable aléatoire $Y$ prend ses valeurs dans $\{1,\frac{1}{2},\frac{1}{3},\frac{1}{4},\frac{1}{5},\frac{1}{6}\}$ et on a clairement que $P(Y=\frac{1}{k})=P(X=k)$. On connait donc la loi de $Y$. Le calcul de l'espérance de $Y$ donne

$$\mathbb{E}(Y)=\sum_{k=1}^{6}\frac{1}{k}P(X=k)=\frac{1}{21}\sum_{k=1}^{6}1=\frac{2}{7}$$

et pour $\mathbb{E}(Y^2)$ on a que

$$\begin{aligned}\mathbb{E}(Y^2)&=P(X=1)+\frac{1}{4}P(X=2)+\frac{1}{9}P(X=3)\\&\quad+\frac{1}{16}P(X=4)++\frac{1}{25}P(X=5)+\frac{1}{36}P(X=6)\\&=\frac{1}{21}\sum_{k=1}^{6}\frac{1}{k}\end{aligned}$$

ce qui donne $\mathbb{E}(Y^2)=\frac{7}{60}$. On trouve ensuite que $\mathrm{Var}(Y)=\frac{7}{60}-\frac{4}{49}=\frac{103}{2.940}$.

## 31. Le cas des lois classiques

C'est un passage obligé dans tout cours de probabilité: on calcule les espérances et les variances des lois classiques rencontrées aux Sections 24, 25 et 26. L'espérance, et donc aussi la variance, ne dépendent que des lois des variables aléatoires considérées (cf. (28.1) et (29.1)). Deux variables aléatoires qui suivent la même loi auront donc les mêmes espérances et variances. Dans ce qui suit, $(\Omega, \mathcal{P}(\Omega), P)$ désigne un espace de probabilité fini et $X : \Omega \to E$ une variable aléatoire réelle. Pour ne pas multiplier les théorèmes, seul le cas de la loi binomiale sera énoncé sous forme de théorème.

**Lemme 31.1** (Espérance et variance pour la loi uniforme)**.** *Si $X$ suit la loi uniforme discrète de paramètre $n$, alors*

$$\mathbb{E}(X) = \frac{n+1}{2} \ \textit{ et } \ Var(X) = \frac{n^2-1}{12} \ .$$

*Démonstration.* On calcule

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{n} kP(X=k) \\ &= \frac{1}{n}\sum_{k=1}^{n} k \\ &= \frac{n+1}{2}\end{aligned}$$

puisque la somme des $n$ premiers entiers vaut

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2} \ . \tag{31.1}$$

On peut démontrer cette identité par récurrence sur $n$ ou en écrivant deux fois la somme considérée sous la forme

$$\begin{array}{ccccc} 1 & 2 & \dots & n-1 & n \\ n & n-1 & \dots & 2 & 1 \end{array}$$

et en additionnant termes à termes pour trouver a chaque fois $n(n+1)$ de sorte que deux fois la somme des $n$ premiers entiers vaut effectivement $n(n+1)$. Au final on a bien démontré la formule annoncée pour l'espérance. De la même façon,

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=1}^{n} k^2 P(X=k) \\ &= \frac{1}{n}\sum_{k=1}^{n} k^2 \\ &= \frac{(n+1)(2n+1)}{6}\end{aligned}$$

car la somme des $n$ premiers carrés vaut

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \tag{31.2}$$

(une formule qui se démontre sans difficulté par récurrence). Avec la formule de Huygens de la Proposition 29.1 on obtient donc

$$\begin{aligned}\mathrm{Var}(X) &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{n+1}{12}\left(2(2n+1) - 3(n+1)\right) \\ &= \frac{(n+1)(n-1)}{12} \\ &= \frac{n^2-1}{12} \ .\end{aligned}$$

On retrouve la formule annoncée pour la variance. Le lemme est démontré. □

**Lemme 31.2** (Espérance et variance pour la loi de Bernoulli)**.** *Si $X$ suit la loi de Bernoulli de paramètre $p$, alors*

$$\mathbb{E}(X) = p \ \textit{et} \ \textit{Var}(X) = p(1-p) \ .$$

*Démonstration.* On a $\mathbb{E}(X) = 1 \times P(X=1) = p$ et $\mathbb{E}(X^2) = 1^2 \times P(X=1) = p$. En particulier on a la formule annoncée pour l'espérance. Avec la formule de Huygens de la Proposition 29.1 on trouve ensuite que $\mathrm{Var}(X) = p - p^2 = p(1-p)$. Le lemme est démontré. □

**Théorème 31.1** (Espérance et variance pour la loi binomiale)**.** *Si $X$ suit la loi binomiale de paramètres $n$ et $p$, alors*

$$\mathbb{E}(X) = np \ \textit{et} \ \textit{Var}(X) = np(1-p) \ .$$

*Démonstration.* On revient à la formule du binôme (7.1) du Théorème 7.1. Cette formule permet en particulier d'écrire que

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

pour tout $x \in \mathbb{R}$ (puisque $1^{n-k} = 1$). En dérivant cette identité et en multipliant par $x$ on obtient que

$$nx(1+x)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} x^k \ . \tag{31.3}$$

On calcule maintenant l'espérance. Grâce à (31.3) on écrit que

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= (1-p)^n \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{p}{1-p}\right)^k \\ &= (1-p)^n n \frac{p}{1-p} \left(1 + \frac{p}{1-p}\right)^{n-1} \\ &= np \ .\end{aligned}$$

On retrouve la formule annoncée pour l'espérance. Pour calculer $\mathbb{E}(X^2)$ on dérive de nouveau. En dérivant (31.3), on obtient que

$$n(1+x)^{n-2}(1+nx) = \sum_{k=0}^{n} k^2 \binom{n}{k} x^{k-1}$$

et en multipliant les deux membres par $x$ on obtient que

$$n(1+x)^{n-2}(x+nx^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} x^{k} \ . \tag{31.4}$$

pour tout $x \in \mathbb{R}$. Grâce à (31.4) on peut alors écrire que

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= (1-p)^n \sum_{k=0}^{n} k^2 \binom{n}{k} \left(\frac{p}{1-p}\right)^k \\
&= (1-p)^n n \left(1+\frac{p}{1-p}\right)^{n-2} \frac{p}{1-p}\left(1+n\frac{p}{1-p}\right) \\
&= np + n(n-1)p^2 \ .
\end{aligned}$$

Avec la formule de Huygens de la Proposition 29.1 on peut maintenant écrire que

$$\begin{aligned}
\mathrm{Var}(X) &= np + n(n-1)p^2 - n^2p^2 \\
&= np(1-p)
\end{aligned}$$

et on retrouve la formule annoncée pour la variance. Le théorème est démontré. □

**Lemme 31.3** (Espérance et variance pour la loi hypergéométrique)**.** *Si $X$ suit la loi hypergéométrique de paramètres $(m_1, m_2, n)$, alors*

$$\mathbb{E}(X) = \frac{m_1 n}{m_1+m_2} \textit{ et } \textit{Var}(X) = \frac{n(m_1+m_2-n)m_1m_2}{(m_1+m_2-1)(m_1+m_2)^2} \ .$$

*Démonstration.* On utilise la relation d'absorption pour les coefficients binomiaux, la seconde formule de (6.2), qui donne que $k\binom{n}{k} = n\binom{n-1}{k-1}$. Avec cette relation on calcule,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{n} k \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}} \\
&= m_1 \sum_{k=1}^{n} \frac{\binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}} \\
&= m_1 \sum_{j=0}^{n-1} \frac{\binom{m_1-1}{j}\binom{m_2}{n-1-j}}{\binom{m_1+m_2}{n}} \\
&= \frac{m_1 n}{m_1+m_2} \sum_{j=0}^{n-1} \frac{\binom{m_1-1}{j}\binom{m_2}{n-1-j}}{\binom{m_1+m_2-1}{n-1}}
\end{aligned}$$

après avoir effectué le changement d'indice $k = 1+j$ et puisque

$$\binom{m_1+m_2}{n} = \frac{m_1+m_2}{n}\binom{m_1+m_2-1}{n-1} \ .$$

La dernière somme revient à sommer les probabilités élémentaires de la loi hypergéométrique de paramètres $(m_1 - 1, m_2, n - 1)$. La somme des probabilités élémentaires vaut 1. On trouve donc bien que

$$\mathbb{E}(X) = \frac{m_1 n}{m_1 + m_2} \ .$$

Le calcul de la variance est fastidieux. On le laisse de côté. L'astuce consiste non pas chercher à calculer $\mathbb{E}(X^2)$ dirèctement, car on se retrouve avec un des $k$ du $k^2$ dont on ne sait pas quoi faire, mais à calculer $\mathbb{E}(X(X-1))$ qui va donner un $k(k-1)$ et permettre d'appliquer deux fois l'astuce précédente basée sur la relation d'absorption. Bien évidemment, si on connait $\mathbb{E}(X(X-1))$ alors on connait $\mathbb{E}(X^2)$ par linéarité de l'espérance puisque $\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X)$ et puisque l'on connait $\mathbb{E}(X)$. Le lemme est démontré. □

On a utilisé dans le calcul ci-dessus que la somme des probabilités élémentaires de la loi hypergéométrique vaut 1. Traduit en terme de relation sur les coefficients binomiaux, cela signifie que

$$\binom{m_1 + m_2}{n} = \sum_{k=0}^{n} \binom{m_1}{k} \binom{m_2}{n-k} \ . \tag{31.5}$$

La relation (31.5) est connue sous les termes *d'identité de Vandermonde.*

## 32. Inégalités de Bienaymé-Tchebychef et Markov

L'inégalité de Bienaymé-Tchebychef montre que la variance mesure effectivement la dispersion d'une variable aléatoire réelle autour de son espérance.

**Théorème 32.1** (Inégalité de Bienaymé-Tchebychef)**.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini et soit* $X : \Omega \to E$ *une variable aléatoire réelle. On a alors que*

$$P\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{1}{a^2} \mathit{Var}(X) \tag{32.1}$$

*pour tout réel* $a > 0$.

*Démonstration.* Soit $a > 0$ donné. Pour tout $y \in \mathbb{R}$, $\mathbb{1}_{[a,+\infty[}(|y|) \leq \frac{1}{a^2} y^2$ puisque $\mathbb{1}_{[a,+\infty[}(|y|) = 0$ si $|y| \leq a$ et $\frac{1}{a^2} y^2 \geq 1$ si $|y| \geq a$. Par suite

$$\mathbb{1}_{\{|X - \mathbb{E}(X)| \geq a\}} \leq \frac{1}{a^2} \left(X - \mathbb{E}(X)\right)^2$$

et ainsi

$$\mathbb{E}\left(\mathbb{1}_{\{|X - \mathbb{E}(X)| \geq a\}}\right) \leq \frac{1}{a^2} \mathrm{Var}(X) \ .$$

Or $\mathbb{1}_{\{|X - \mathbb{E}(X)| \geq a\}}$ est une variable aléatoire qui suit une loi de Bernoulli de paramètre $p = P\left(|X - \mathbb{E}(X)| \geq a\right)$ (cf. Théorème 24.1 et la discussion après ce théorème) et donc (cf. Section 31),

$$\mathbb{E}\left(\mathbb{1}_{\{|X - \mathbb{E}(X)| \geq a\}}\right) = P\left(|X - \mathbb{E}(X)| \geq a\right) \ .$$

D'où l'inégalité (32.2). Le théorème est démontré. □

Posons par exemple $a = 10\sigma(X)$, où $\sigma(X)$ est l'écart type défini à la Définition 29.1. On trouve alors que la probabilité que $X$ dévie de plus de $10\sigma(X)$ de sa moyenne $\mathbb{E}(X)$ est inférieure à 1%. En prenant $a = 3,17\sigma(X)$ on trouve que la probabilité que $X$ dévie de plus de $3,17\sigma(X)$ de sa moyenne $\mathbb{E}(X)$ est inférieure à 10%. On utilise ici le langage des pourcentages qui est plus parlant dans le langage

courant, mais bien sur 1% signifie 0,01 et 10% signifie 0,1 pour les probabilités selon leur définition première (règles de trois qui ramène 100, langage des pourcentages, à 1 pour les probabilités. Par exemple: à 10 correspond 100, à $x$ correspond 1, donc $x = 0,1$). Un autre exemple d'application est donné par le cas suivant: on jette un dé équilibré 3600 fois et on note $X$ la variable aléatoire comptant le nombre de fois où le 4 est apparu. On cherche a minorer $P(480 < X < 720)$. On remarque pour commencer que l'on est en train de compter un nombre de succès avec une probabilité $1/6$ de succès à chaque lancé. Donc $X$ suit la loi binomiale de paramètres 3600 et $1/6$. Par suite, en raison du Théorème 31.1, $\mathbb{E}(X) = 3600/6 = 600$ et

$$\mathrm{Var}(X) = \frac{3600}{6}\left(1 - \frac{1}{6}\right) = 500 \ .$$

On remarque ensuite que

$$\begin{aligned} 480 < X < 720 &\Leftrightarrow 480 - 600 < X - 600 < 720 - 600 \\ &\Leftrightarrow -120 < X - 600 < 120 \\ &\Leftrightarrow |X - \mathbb{E}(X)| < 120 \ . \end{aligned}$$

Par inégalité de Bienaymé-Tchebychef,

$$P(|X - \mathbb{E}(X)| \geq 120) \leq \frac{500}{120^2} = \frac{5}{144} \ .$$

Par suite,

$$\begin{aligned} P(480 < X < 720) &= P(|X - \mathbb{E}(X)| < 120) \\ &= 1 - P(|X - \mathbb{E}(X)| \geq 120) \\ &\geq 1 - \frac{5}{144} = \frac{139}{144} \end{aligned}$$

et on trouve donc que la probabilité que $X$ se situe entre 480 et 720 est supérieure à $0,96$. Le résultat qui suit constitue une amélioration de l'inégalité de Bienaymé-Tchebychef pour les $0 < a \ll 1$ lorsque $\mathrm{Var}(X) > 0$.

**Théorème 32.2** (Inégalité de Cantelli)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soit $X : \Omega \to E$ une variable aléatoire réelle. On a alors que*

$$P\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{2\,Var(X)}{a^2 + Var(X)} \tag{32.2}$$

*pour tout réel $a > 0$.*

*Démonstration.* La preuve procède en plusieurs étapes. Soit $a > 0$ fixé. Pour commencer on démontre que

$$P\left(X - \mathbb{E}(X) \geq a\right) \leq \frac{t^2 + \mathrm{Var}(X)}{(a+t)^2} \tag{32.3}$$

pour tout $t \geq 0$. Pour démontrer (32.3) on fixe $t \geq 0$ et on pose $Y = X - \mathbb{E}(X) + t$. On a alors

$$X - \mathbb{E}(X) \geq a \ \Rightarrow \ Y \geq a + t \ \Rightarrow \ Y^2 \geq (a+t)^2$$

puisque $a + t \geq 0$. Donc

$$\begin{aligned} P(X - \mathbb{E}(X) \geq a) &\leq P(Y^2 \geq (a+t)^2) \\ &\leq \frac{\mathbb{E}(Y^2)}{(a+t)^2} \end{aligned}$$

par inégalité de Markov, le Théorème 32.3 ci-dessous, appliquée à $Y^2$, et il reste à remarquer que

$$\begin{aligned}\mathbb{E}(Y^2) &= \mathbb{E}\left((X-\mathbb{E}(X)+t)^2\right)\\ &= \mathbb{E}\left((X-\mathbb{E}(X))^2+2t(X-\mathbb{E}(X))+t^2\right)\\ &= \mathbb{E}\left((X-\mathbb{E}(X))^2\right)+2t\mathbb{E}\left(X-\mathbb{E}(X)\right)+t^2\\ &= \mathrm{Var}(X)+t^2\end{aligned}$$

puisque $\mathbb{E}\left(X-\mathbb{E}(X)\right)=0$. On a donc bien démontré (32.3). On montre maintenant que

$$P\left(X-\mathbb{E}(X)\geq a\right)\leq \frac{\mathrm{Var}(X)}{a^2+\mathrm{Var}(X)} \tag{32.4}$$

Pour cela il suffit de trouver le minimum en $t$ de la fonction $t\to\frac{t^2+\mathrm{Var}(X)}{(a+t)^2}$. Si $\Phi$ est cette fonction,

$$\begin{aligned}\Phi'(t) &= \frac{2t(a+t)^2-2(t^2+\mathrm{Var}(X))(a+t)}{(a+t)^4}\\ &= \frac{2\left(at-\mathrm{Var}(X)\right)}{(a+t)^3}\end{aligned}$$

et on voit ainsi facilement que $\Phi$ a un minimum en $t_0=\frac{\mathrm{Var}(X)}{a}$. Or

$$\Phi(t_0)=\frac{\mathrm{Var}(X)}{a^2+\mathrm{Var}(X)}$$

et on obtient ainsi (32.4) à partir de (32.3). On a que $X-\mathbb{E}(X)\leq -a$ équivaut à $Y-\mathbb{E}(Y)\geq -a$, où $Y=-X$, et donc, avec (32.4) appliqué à $Y$,

$$\begin{aligned}P\left(X-\mathbb{E}(X)\leq -a\right) &= P(Y-\mathbb{E}(Y)\geq a)\\ &\leq \frac{\mathrm{Var}(Y)}{a^2+\mathrm{Var}(Y)}\\ &\leq \frac{\mathrm{Var}(X)}{a^2+\mathrm{Var}(X)}\end{aligned}$$

puisque $\mathrm{Var}(Y)=\mathrm{Var}(X)$ en raison, par exemple, du Théorème 29.2. Reste maintenant à remarquer que

$$\{|X-\mathbb{E}(X)|\geq a\}=\{X-\mathbb{E}(X)\leq -a\}\bigcup\{X-\mathbb{E}(X)\geq a\}$$

et que la réunion est disjointe pour obtenir l'inégalité de Cantelli à partir de (32.4) et de ce qui vient d'étre dit. □

On remarque que $\frac{2\mathrm{Var}(X)}{a^2+\mathrm{Var}(X)}<\frac{1}{a^2}\mathrm{Var}(X)\quad\Leftrightarrow\quad \mathrm{Var}(X)>a^2$ et donc que l'inégalité de Cantelli améliore Binaymé-Tchebychef pour les $a>0$ petits lorsque $\mathrm{Var}(X)>0$. On traite maintenant de l'inégalité de Markov qui constitue une des inégalités les plus célèbres de la théorie des probabilités.

**Théorème 32.3** (Inégalité de Markov)**.** *Soit $(\Omega,\mathcal{P}(\Omega),P)$ un espace de probabilité fini et soit $X:\Omega\to E$ une variable aléatoire réelle à valeurs positives ou nulles. Alors*

$$P(X\geq x)\leq\frac{1}{x}\mathbb{E}(X) \tag{32.5}$$

*pour tout* $x > 0$.

*Démonstration.* Soit $x > 0$. On écrit que $1 = \mathbb{1}_{\{X \geq x\}} + \mathbb{1}_{\{X < x\}}$. Alors

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X\mathbb{1}_{\{X \geq x\}}) + \mathbb{E}(X\mathbb{1}_{\{X < x\}}) \\ &\geq \mathbb{E}(X\mathbb{1}_{\{X \geq x\}})\end{aligned}$$

puisque $X$ est à valeurs positives ou nulles. En remarquant que

$$\mathbb{E}(X\mathbb{1}_{\{X \geq x\}}) \geq xP(X \geq x) \ ,$$

on obtient l'inégalité de Markov (32.5). Le théorème est démontré. □

Une variante autour de l'inégalité de Markov est donnée par le résultat suivant.

**Théorème 32.4.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini et soit* $X : \Omega \to E$ *une variable aléatoire réelle à valeurs positives ou nulles. Soit* $f : \mathbb{R}^+ \to \mathbb{R}^{+\star}$ *une application strictement croissante de* $\mathbb{R}^+$ *dans* $\mathbb{R}^{+\star}$. *Alors*

$$P(x \geq x) \leq \frac{\mathbb{E}\left(f(X)\right)}{f(x)}$$

*pour tout* $x \geq 0$.

*Démonstration.* Soit $x \geq 0$. Pour $\omega \in \Omega$,

$$X(\omega) \geq x \ \Leftrightarrow \ f\left(X(\omega)\right) \geq f(x)$$

par stricte croissance de $f$, et $f(x) > 0$ puisque $f$ est à valeurs dans $\mathbb{R}^{+\star}$. Soit $Y = f(X)$. En appliquant Markov a $Y$,

$$\begin{aligned}P(X \geq x) &= P(Y \geq f(x)) \\ &\leq \frac{1}{f(x)}\mathbb{E}(Y)\end{aligned}$$

ce qui constitue exactement l'inégalité voulue. Le théorème est démontré. □

## 33. Covariance et corrélation

On parle de covariance pour un couple de variables aléatoires.

**Définition 33.1** (Covariance)**.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini et soient* $X : \Omega \to E$ *et* $Y : \Omega \to E'$ *deux variables aléatoires réelles. On appelle covariance de* $X$ *et* $Y$ *le nombre*

$$\begin{aligned}Cov(X, Y) &= \mathbb{E}\left[(X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\right] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \ .\end{aligned} \tag{33.1}$$

*Lorsque* $Var(X) > 0$ *et* $Var(Y) > 0$ *le coefficient de correlation* $Cor(X, Y)$ *de* $X$ *et* $Y$ *est donné par la formule* $Cov(X, Y) = Cor(X, Y)\sqrt{Var(X)\,Var(Y)}$.

Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. L'application définie sur les couples de variables aléatoires réelles de $(\Omega, \mathcal{P}(\Omega), P)$ par $\Phi : (X, Y) \to \mathbb{E}(XY)$ est une forme bilinéaire symétrique positive (application à valeurs dans $\mathbb{R}$, linéaire en $X$, linéaire en $Y$, symétrique en $X$ et $Y$ et toujours positive ou nulle sur les couples $(X, X)$). De telles applications vérifient l'inégalité dite de Cauchy-Schwarz:

$$|\Phi(X, Y)| \leq \sqrt{\Phi(X, X)}\sqrt{\Phi(Y, Y)} \tag{33.2}$$

pour tous $X, Y$. En effet, pour $X, Y$ quelconques, $\Phi(X + xY, X + xY) \geq 0$ pour tout réel $x$. Or, par bilinéarité,

$$\Phi(X + xY, X + xY) = \Phi(Y, Y)x^2 + 2\Phi(X, Y)x + \Phi(X, X) \ .$$

Si $\Phi(Y, Y) = 0$ on a là une fonction affine que ne change pas de signe, donc forcément constante et $\Phi(X, Y) = 0$ de sorte que l'inégalité de Cauchy-Schwarz est trivialement vraie. Si $\Phi(Y, Y) \neq 0$, on a là un polynôme du second degré en $x$. Un polynôme du second degré qui a deux racines change forcément de signe (entre ses racines). Donc ce polynôme a au plus une racine et son discriminant $\Delta$ est forcément négatif ou nul, ce qui revient à dire que (33.2) a lieu.

**Proposition 33.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et $X : \Omega \to E$, $Y : \Omega \to E'$ deux variables aléatoires réelles telles que $Var(X) > 0$ et $Var(Y) > 0$. On a $Cor(X, Y) \in [-1, +1]$.*

*Démonstration.* L'inégalité de Cauchy-Schwarz (33.2) pour $\Phi(X, Y) = \mathbb{E}(XY)$ appliquée à $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ donne que $|\mathrm{Cor}(X, Y)| \leq 1$. La proposition suit. □

Le lien entre variance et covariance suit de la proposition suivante. C'est bien sûr une relation du type $(x+y)^2 = x^2+y^2+2xy$. La covariance est bilinéaire (linéaire en $X$ à $Y$ fixée, et linéaire en $Y$ à $X$ fixée), elle est symétrique et $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ pour toute variable aléatoire $X$.

**Proposition 33.2.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. On a*

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y) \tag{33.3}$$

*pour toutes variables aléatoires réelles $X, Y$ sur $(\Omega, \mathcal{P}(\Omega), P)$.*

*Démonstration.* On développe en utilisant la linéarité de l'espérance. Soient donc $X : \Omega \to E$ et $Y : \Omega \to E'$ deux variables aléatoires réelles. Alors

$$\begin{aligned}
\mathrm{Var}(X + Y) &= \mathbb{E}(X^2 + Y^2 + 2XY) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\left(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) \ .
\end{aligned}$$

La proposition est démontrée. □

La formule (33.3) se généralise à la somme de plusieurs variables aléatoires:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) \tag{33.4}$$

pour tout $n \geq 2$ et toutes variables aléatoires réelles $X_1, \ldots, X_n$ sur $(\Omega, \mathcal{P}(\Omega), P)$. La preuve de (33.4) s'obtient facilement par récurrence avec (33.3) et la bilinéarité de la covariance.

## 34. Familles de variables aléatoires

On aborde dans cette section la notion d'indépendances de variables aléatoires et ses relations avec la variance et la covariance. La notion de variables aléatoires indépendantes renvoie ici aux notions d'événements mutuellement indépendants et d'événements deux à deux indépendants.

**Définition 34.1** (Variables aléatoires indépendantes)**.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini. Soient* $n \geq 2$ *et* $X_i : \Omega \to E_i$ *des variables aléatoires sur* $(\Omega, \mathcal{P}(\Omega), P)$*,* $i = 1, \dots, n$*. Alors:* **(1)** *On dit que les* $X_i$ *sont indépendantes si*

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

*pour tous* $x_i \in E_i$*,* $i = 1, \dots, n$*. Et:* **(2)** *On dit que les* $X_i$ *sont deux à deux indépendantes si pour tous* $i \neq j$*,* $X_i$ *et* $X_j$ *sont indépendantes.*

Plusieurs propriétés sont à démontrer. La première a trait à la notion (1) d'indépendance. Elle illustre le fait que si $X_1, \dots, X_n$ sont indépendantes, alors toute sous famille $X_{i_1}, \dots, X_{i_p}$ des $X_i$ est encore indépendante. En particulier, pour tout $(x_1, \dots, x_n) \in E_1 \times \cdots \times E_n$, les événements $\{X_1 = x_1\}, \dots, \{X_n = x_n\}$ sont mutuellement indépendants, et l'indépendance entraîne l'indépendance deux à deux.

**Proposition 34.1.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini. Soient* $n \geq 2$ *et* $X_i : \Omega \to E_i$ *des variables aléatoires indépendantes sur* $(\Omega, \mathcal{P}(\Omega), P)$*,* $i = 1, \dots, n$*. Alors pour toute sous famille* $i_1, \dots, i_p$ *d'indices dans* $\{1, \dots, n\}$*,* $1 \leq p \leq n$*, les* $X_{i_1}, \dots, X_{i_p}$ *sont encore indépendantes. En particulier, l'indépendance entraîne l'indépendance deux à deux.*

*Démonstration.* Soient $i_1, \dots, i_p$ fixés et $x_{i_1}, \dots, x_{i_p}$ donnés quelconques. On note $j_1, \dots, j_q$, $q = n - p$, les indices qui manquent pour former $\{1, \dots, n\}$. On a alors

$$\bigcap_{k=1}^{p} \{X_{i_k} = x_{i_k}\} = \bigcup_{\{x_{j_\ell} \in E_{j_\ell}, \ell = 1, \dots, q\}} \{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\} \ ,$$

où $x_i = x_{i_k}$ si $i = i_k$ et $x_i = x_{j_\ell}$ si $i = j_\ell$. La réunion est disjointe et par additivité (cf. Proposition 12.2) on obtient que

$$\begin{aligned} P(\cap_{k=1}^{p} \{X_{i_k} = x_{i_k}\}) &= \sum_{\{x_{j_\ell} \in E_{j_\ell}, \ell = 1, \dots, q\}} P(X_1 = x_1) \times \cdots \times P(X_n = x_n) \\ &= P(X_{i_1} = x_{i_1}) \times \cdots \times P(X_{i_p} = x_{i_p}) \end{aligned}$$

puisque pour tout $\ell = 1, \dots, q$, la somme des probabilités élémentaires valant 1,

$$\sum_{\{x_{j_\ell} \in E_{j_\ell}\}} P(X_{j_\ell} = x_{j_\ell}) = 1 \ .$$

Les $x_{i_1}, \dots, x_{i_p}$ sont quelconques. Donc les $X_{i_1}, \dots, X_{i_p}$ sont encore indépendantes. Comme la sous famille $i_1, \dots, i_p$ d'indices est quelconque dans ce raisonnement, la proposition est démontrée. □

La proposition qui suit établit un lien entre indépendance et espérance.

**Théorème 34.1.** *Soit* $(\Omega, \mathcal{P}(\Omega), P)$ *un espace de probabilité fini. Soient* $n \geq 2$ *et* $X_i : \Omega \to E_i$ *des variables aléatoires réelles sur* $(\Omega, \mathcal{P}(\Omega), P)$*. Ces variables sont indépendantes si et seulement si pour toutes fonctions réelles* $f_1, \dots, f_n$*,*

$$\mathbb{E}\left(\prod_{i=1}^{n} f_i(X_i)\right) = \prod_{i=1}^{n} \mathbb{E}\left(f_i(X_i)\right) \ . \tag{34.1}$$

*Démonstration.* (1) On suppose pour commencer que les $X_i$ sont indépendantes et on montre la relation (34.1) du théorème. Soient $f_i : E_i \to \mathbb{R}$ des fonctions. On note $X = (X_1, \dots, X_n)$ la variable aléatoire à valeurs dans $E = E_1 \times \cdots \times E_n$ et $f : E \to \mathbb{R}$ l'application $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i$. Par transport de l'espérance (cf. Proposition 28.3), et puisque les variables aléatoires sont indépendantes, on a

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) &= \mathbb{E}\left(f(X)\right) \\
&= \sum_{x \in E} f(x) P(X = x) \\
&= \sum_{x_i \in E_i} \left(\prod_{i=1}^n f_i(x_i)\right) P\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) \\
&= \sum_{x_i \in E_i} \prod_{i=1}^n f_i(x_i) P(X_i = x_i) \\
&= \prod_{i=1}^n \sum_{x_i \in E_i} f_i(x_i) P(X_i = x_i) \\
&= \prod_{i=1}^n \mathbb{E}\left(f_i(X_i)\right) \ .
\end{aligned}$$

La relation (34.1) est démontrée. (2) A l'inverse on suppose que pour toutes fonctions réelles $f_1, \dots, f_n$, (34.1) est vraie. Soient $x_1, \dots, x_n$ donnés quelconques. On pose $f_i = \mathbb{1}_{\{x_i\}}$. Alors, avec (34.1),

$$P\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) = \mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(f_i(X_i)\right) = \prod_{i=1}^n P(X_i = x_i)$$

et les variables aléatoires $X_1, \dots, X_n$ sont donc bien indépendantes. Le théorème est démontré. □

On regarde maintenant les relations entre indépendance et covariance.

**Théorème 34.2.** *Soient $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et $X : \Omega \to E$, $Y : \Omega \to E'$ deux variables aléatoires réelles. Si $X$ et $Y$ sont indépendantes, alors $Cov(X, Y) = 0$. Par contre la réciproque est fausse en général.*

*Démonstration.* Supposons que $X$ et $Y$ sont indépendantes. D'après le Théorème 34.1 avec $f_1 = f_2 = \mathrm{id}$ (où id est l'identité de $\mathbb{R}$), $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Donc $\mathrm{Cov}(X, Y) = 0$. Pour montrer que la réciproque est fausse en général, considérons $X_1$ et $X_2$ deux variables aléatoires indépendantes qui suivent une loi de Bernoulli de paramètre $\frac{1}{2}$. On pose $X = X_1 + X_2$ et $Y = |X_1 - X_2|$. Par linéarité de l'espérance (cf. Proposition 28.1), et avec les résultats de la Section 31, $\mathbb{E}(X) = 1$. La variable aléatoire $Y$ prend les valeurs 0 et 1 et

$$\mathbb{E}(Y) = P(Y = 1) = 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$$

car $\{Y = 1\} = \big(\{X_1 = 1\} \cap \{X_2 = 0\}\big) \bigcup \big(\{X_1 = 0\} \cap \{X_2 = 1\}\big)$, la réunion est disjointe et $X_1$, $X_2$ sont indépendantes. La variable aléatoire $XY$, elle, prend les

valeurs 0, 1 et 2 car $X$ prend les valeurs 0, 1 et 2. On a

$$\begin{aligned}\{XY = 1\} &= \{X = 1\} \cap \{Y = 1\} \\ &= \big(\{X_1 = 1\} \cap \{X_2 = 0\}\big) \bigcup \big(\{X_1 = 0\} \cap \{X_2 = 1\}\big) \\ &= \{Y = 1\}\end{aligned}$$

et $\{XY = 2\} = \{X = 2\} \cap \{Y = 1\} = \emptyset$ puisque $X = 2 \Leftrightarrow X_1 = X_2 = 1$ et $X_1 = 1, X_2 = 1 \Rightarrow Y = 0$. Par suite,

$$\begin{aligned}\mathbb{E}(XY) &= P(XY = 1) + 2P(XY = 2) \\ &= P(Y = 1) \\ &= \frac{1}{2}\end{aligned}$$

On en déduit que $\mathrm{Cov}(X, Y) = \frac{1}{2} - 1 \times \frac{1}{2} = 0$. Reste à montrer que $X$ et $Y$ ne sont pourtant pas indépendantes. Or $\{X = 1\} \cap \{Y = 1\} = \{Y = 1\}$ (cf. ci-dessus) et, avec le même raisonnement, $\{X = 1\} = \{Y = 1\}$. Si $X$ et $Y$ étaient indépendantes on aurait $P(\{X = 1\} \cap \{Y = 1\}) = P(X = 1)P(Y = 1)$ et donc $P(X = 1) = P(Y = 1) = 0$ ou 1. Or $P(X = 1) = P(Y = 1) = \frac{1}{2}$. Les variables aléatoires $X$ et $Y$ ne sont donc pas indépendantes. Le théorème est démontré. □

Les deux résultats qui suivent sont des conséquences immédiates de ce qui précède. Le premier de ces deux résultats donne une formule pour la variance d'une somme de variables aléatoires deux à deux indépendantes.

**Théorème 34.3.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soient $n \geq 2$ et $X_i : \Omega \to E_i$ des variables aléatoires réelles deux à deux indépendantes sur $(\Omega, \mathcal{P}(\Omega), P)$. Alors $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$.*

*Démonstration.* Le théorème suit dirèctement de (33.4) et du Théorème 34.2. □

On termine cette section avec le résultat suivant.

**Théorème 34.4.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soient $n \geq 2$ et $X_i : \Omega \to E_i$ des variables aléatoires réelles indépendantes et qui suivent une même loi de Bernoulli de paramètre $p$ sur $(\Omega, \mathcal{P}(\Omega), P)$. Alors $X = \sum_{i=1}^n X_i$ suit une loi binomiale de paramètres $n$ et $p$.*

*Démonstration.* On pose $A_i = \{X_i = 1\}$. Les $A_i$ sont indépendants puisque les $X_i$ sont indépendantes. On a $X = \sum_{i=1}^n \mathbb{1}_{A_i}$ et les $P(A_i)$ sont tous égaux à $p$. Le résultat suit alors du Théorème 25.1. □

## 35. Quelques exercices 7

**EXERCICE 27*:** Soient $X, Y$ deux variables aléatoires indépendantes qui suivent la loi uniforme discrète de paramètre $n$. Déterminer $P(X = Y)$, $P(X \geq Y)$ et la loi de $X + Y$.

CORRECTION: On a $\{X = Y\} = \bigcup_{k=1}^n \{X = k\} \cap \{Y = k\}$ et les ensembles dans l'union sont deux à deux disjoints. Donc, comme les variables aléatoires sont indépendantes, $P(X = Y) = \sum_{k=1}^n P(X = k)P(Y = k) = \frac{1}{n}$ puisque pour tout $k$, $P(X = k) = P(Y = k) = \frac{1}{n}$. Pour calculer $P(X \geq Y)$ on peut remarquer que par

symétrie, $P(X \geq Y) = P(Y \geq X)$. Par ailleurs, si $\Omega$ est l'univers des possibles, alors $\Omega = \{X \geq Y\} \cup \{Y \geq X\}$ et donc (cf. Théorème 12.1),

$$\begin{aligned} 1 &= P(\{X \geq Y\} \cup \{Y \geq X\}) \\ &= P(X \geq Y) + P(Y \geq X) - P(X = Y) \\ &= 2P(X \geq Y) - P(X = Y) \end{aligned}$$

puisque $\{X \geq Y\} \cap \{Y \geq X\} = \{X = Y\}$. On en déduit que $P(X \geq Y) = \frac{1}{2}(1 + \frac{1}{n})$. Reste à déterminer la loi de $X + Y$. La variable aléatoire $X + Y$ est à valeurs dans $[\![2, 2n]\!]$. Pour $k \in [\![2, 2n]\!]$,

$$\{X + Y = k\} = \bigcup_{i=1}^{k-1} \{X = i\} \cap \{Y = k - i\}$$

et les ensembles dans l'union sont deux à deux disjoints. Toujours par indépendance de $X$ et $Y$ on peut alors écrire que

$$P(X + Y = k) = \sum_{i=1}^{k-1} P(X = i)P(Y = k - i) \ .$$

Si $k > n$, $P(X = i) = 0$ pour $n < i \leq k - 1$ et on a aussi $P(Y = k - i) = 0$ si $k - i > n$. On veut donc $i \leq n$ et $k - i \leq n$ soit $k - n \leq i \leq n$. Donc, si $k > n$, $P(X + Y = k) = \sum_{k-n}^{n} \frac{1}{n^2} = \frac{2n+1-k}{n^2}$. Si par contre $k \leq n$ alors $k - i \leq n$ pour tout $i \geq 1$ et donc $P(X + Y = k) = \sum_{i=1}^{k-1} \frac{1}{n^2} = \frac{k-1}{n^2}$.

**EXERCICE 28$^\star$:** On lance deux dés équilibrés et on note $U_1$, $U_2$ les variables aléatoires correspondant aux résultats obtenus. On note $X = \min(U_1, U_2)$ et aussi $Y = \max(U_1, U_2)$. Déterminer la loi de $X$ et calculer $\mathbb{E}(X)$. Calculer $\mathbb{E}(X + Y)$ puis $\mathbb{E}(Y)$. Déterminer $\mathbb{E}(XY)$ puis $\mathrm{Cov}(X, Y)$. Les variables aléatoires $X$ et $Y$ sont elles indépendantes ?

CORRECTION: Les variables aléatoires $X$ et $Y$ sont toutes deux à valeurs dans $[\![1, 6]\!]$. L'univers des possibles $\Omega$ est l'ensemble de tous les couples possibles, $\sharp\Omega = 36$. Par équiprobabilité, en listant les cas, on calcule

$$\begin{aligned} P(X = 1) &= \frac{11}{36} \ , \ P(X = 2) = \frac{9}{36} \ , \ P(X = 3) = \frac{7}{36} \ , \\ P(X = 4) &= \frac{5}{36} \ , \ P(X = 5) = \frac{3}{36} \ , \ P(X = 6) = \frac{1}{36} \ . \end{aligned}$$

Il s'ensuit que $\mathbb{E}(X) = \sum_{k=1}^{6} kP(X = k) = \frac{91}{36}$. On remarque maintenant que $X + Y = U_1 + U_2$ car $\min(x, y) + \max(x, y) = x + y$ pour tous $x, y \in \mathbb{R}$. Donc $\mathbb{E}(X + Y) = \mathbb{E}(U_1 + U_2) = \mathbb{E}(U_1) + \mathbb{E}(U_2)$, et comme $U_1$ et $U_2$ suivent la loi uniforme discrète de paramètre 6, on obtient que (cf. Lemma 31.1): $\mathbb{E}(X + Y) = 7$. Par linéarité de l'espérance (cf. Proposition 28.1), $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ et donc $\mathbb{E}(Y) = 7 - \frac{91}{36} = \frac{161}{36}$. On cherche maintenant $\mathbb{E}(XY)$. On remarque que $XY = U_1U_2$ car $\min(x, y) \times \max(x, y) = xy$ pour tous $x, y \in \mathbb{R}$. Avec le Théorème 34.1, puisque $U_1$ et $U_2$ sont indépendantes, $\mathbb{E}(U_1U_2) = \mathbb{E}(U_1)\mathbb{E}(U_2) = \frac{49}{4}$. Donc

$\mathbb{E}(XY) = \frac{49}{4}$. Reste à calculer la covariance de $X$ et $Y$. On a

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\\ &= \frac{49}{4} - \frac{91}{36} \times \frac{161}{36}\\ &= \frac{1.225}{1.296}\ .\end{aligned}$$

On a $\mathrm{Cov}(X,Y) \neq 0$. D'après le Théorème 34.2 les variables aléatoires $X$ et $Y$ ne sont pas indépendantes, car sinon on aurait $\mathrm{Cov}(X,Y) = 0$.

**EXERCICE 29:** Soient $X_1, \dots, X_n$ des variables aléatoires réduites sur un même espace. On suppose que les $X_i$ sont deux à deux indépendantes et que les $X_i$ suivent une même loi. On note $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $E = \mathbb{E}(X_1)$. Montrer que pour tout $\varepsilon > 0$, $P\left(|\overline{X}_n - E| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}$.

CORRECTION: Il s'agit là d'une application assez dirècte de l'inégalité de Bienaymé-Tchebychef du Théorème 32.1. On a

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n}$$

avec la Proposition 29.2, le Lemme 31.2 et le Théorème 34.3, puisque les variables aléatoires $X_n$ sont deux à deux indépendantes et réduites de sorte que $\mathrm{Var}(X_i) = 1$ pour tout $i$. Par linéarité de l'espérance (cf. Proposition 28.1), $\mathbb{E}(\overline{X}_n) = E$ puisque les $X_i$ suivent une même loi. Soit $\varepsilon > 0$ fixé quelconque. L'inégalité de Bienaymé-Tchebychef du Théorème 32.1 permet d'écrire que

$$P\left(|\overline{X}_n - E| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}$$

et on obtient donc le résultat demandé.

## 36. Loi faible des grands nombres

La loi faible des grands nombres illustre l'idée forte que lorsque l'on répète une expérience aléatoire un grand nombre de fois, dans les mêmes conditions, de façon indépendantes les unes des autres, la moyenne des résultats obtenus tend vers une valeur fixée. On parle de lois parce que ces résultats ont été utilisés avant leur démonstration complète. Dans le cas fini la loi faible des grands nombres nécessite quelques manipulations pour être énoncée rigoureusement. On considère $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini. Soit $X : \Omega \to E$ une variable aléatoire réelle modélisant une expérience aléatoire. On répète cette expérience $n$ fois de façon indépendante. On passe alors sur une variable aléatoire

$$X^n : \Omega^n \to E^n$$

donnée par $X^n(\omega_1, \dots, \omega_n) = (X(\omega_1), \dots, X(\omega_n))$. On note $X_i = p_i \circ X^n : \Omega^n \to E$ où $p_i : \mathbb{R}^n \to \mathbb{R}$ est la ième projection. On note $P_n$ la probabilité produit sur $(\Omega^n, \mathcal{P}(\Omega^n))$ donnée par la loi de probabilité

$$P_n(\{(\omega_1, \dots, \omega_n)\}) = \prod_{i=1}^n P(\{\omega_i\})\ . \tag{36.1}$$

Les $X_1, \dots, X_n$ sont deux à deux indépendantes et ont même loi, en particulier elles ont donc même espérance et même variance. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ . \tag{36.2}$$

qui est une variable aléatoire sur $\Omega^n$. On dit que $\overline{X}_n$ est la moyenne empirique des $X_1, \dots, X_n$. On remarque que

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i)$$

en raison du théorème 29.2 et que, en vertue du Théorème 34.3,

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = n\mathrm{Var}(X_1)$$

puisque les $X_i$ ont même loi et sont indépendantes. Dans le cas modèle où l'on répète une même expérience, on a $\mathrm{Var}(X_1) = \mathrm{Var}(X)$ et $\mathbb{E}(X_1) = \mathbb{E}(X)$ puisque si $E = \{x_1, \dots, x_k\}$ alors

$$\begin{aligned}\mathbb{E}(X_1) &= \sum_{i=1}^{k} x_i P(X = x_i) \\ &= \mathbb{E}(X)\end{aligned}$$

et

$$\begin{aligned}\mathrm{Var}(X_1) &= \sum_{i=1}^{k} x_i^2 P(X = x_i) - \left(\sum_{i=1}^{k} x_i P(X = x_i)\right)^2 \\ &= \mathrm{Var}(X)\end{aligned}$$

en remarquant que $\{X_1 = x_i\} = \{X = x_i\} \times \Omega \times \cdots \times \Omega$ de sorte que

$$\begin{aligned}P(X_1 = x_i) &= \sum_{\{(\omega_1,\dots,\omega_n)/X(\omega_1)=x_i\}} P(\{\omega_1\})P(\{\omega_2\})\dots P(\{\omega_n\}) \\ &= P(X = x_i) \times 1 \times \cdots \times 1 \\ &= P(X = x_i) \ .\end{aligned}$$

On peut alors énoncer la loi faible des grands nombres dans le cas discret fini qui modélise donc cette situation de repétition d'une même expérience $n$ fois avec $n \to +\infty$.

**Théorème 36.1** (Loi faible des grands nombres)**.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini et soient $\mu \in \mathbb{R}$ et $\nu \in \mathbb{R}^+$ deux réels. Pour tout $n \geq 1$ on se donne $X_1, \dots, X_n$ des variables aléatoires réelles deux à deux indépendantes sur $(\Omega^n, \mathcal{P}(\Omega^n), P_n)$, où $P_n$ est la loi produit donnée par* (36.1)*. On suppose que les $X_1, \dots, X_n$ ont une même loi d'espérance $\mu$ et de variance $\nu$ (indépendantes de $n$). Alors la suite des moyennes empiriques $\overline{X}_n$ données par* (36.2) *converge en probabilité vers $\mathbb{E}(X_1)$ au sens où*

$$\lim_{n\to+\infty} P_n\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) = 0 \tag{36.3}$$

*pour tout $\varepsilon > 0$. La moyenne empirique converge en probabilité vers $\mu = \mathbb{E}(X_1)$ dès lors que les $X_n$ sont indépendantes et de même loi.*

*Démonstration.* Le théorème est une conséquence dirècte de l'inégalité de Bienaymé-Tchebychef et des propriétés de l'espérance. Par linéarité de l'espérance, cf. Proposition 28.1, et puisque les $X_n$ suivent toutes la même loi,

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1) = \mu \tag{36.4}$$

pour tout $n$. Soit $\varepsilon > 0$ fixé. D'après l'inégalité de Bienaymé-Tchebychef, cf. Théorème 32.1, on a alors que

$$P_n\left(\left|\overline{X}_n - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathrm{Var}(\overline{X}_n) \tag{36.5}$$

pour tout $n$. En vertue de la Proposition 29.2 et du Théorème 34.3,

$$\begin{aligned}\mathrm{Var}(\overline{X}_n) &= \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^n X_i)\\ &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i)\\ &= \frac{1}{n}\nu\end{aligned} \tag{36.6}$$

puisque les $X_n$ suivent tous la même loi de variance $\nu$. En combinant (36.4), (36.5) et (36.6) on obtient (36.3). Le théorème est démontré. □

Le fait que l'on mesure par rapport à une probabilité $P_n$ qui dépend de $n$ est déprimant. Mais en fait on peut s'en débarrasser. Vous verrez en L3 qu'on peut construire un espace $\Omega^\infty$ et une probabilité $P_\infty$ sur $\Omega^\infty$ tels que $\Omega^n \subset \Omega^\infty$ pour tout $n$ et $P_\infty = P_n$ sur $\Omega^n$ pour tout $n$. Cela permet de remplacer la probabilité $P_n$ par une probabilité fixe $P_\infty$. Vous construirez un espace universel en quelque sorte.

Une intuition sur comment se débarasser du $P_n$ peut être donnée par la remarque suivante. Pour $m \geq n$ notons $P_m^n : \Omega^m \to \Omega^n$ l'application d'oubli $P_m^n(\omega_1, \ldots, \omega_m) = (\omega_1, \ldots, \omega_n)$. Si $Y_n$ est une variable aléatoire sur $\Omega_n$, $Y_n \circ P_m^n$ est une variable aléatoire sur $\Omega^m$ et $P_m(Y_n \circ P_m^n = k) = P_n(Y_n = k)$ pour tout $k$. On pourra écrire $P_m(Yn = k) = P_n(Y_n = k)$ pour tout $k$, puisqu'en fait $Y_n \circ P_m^n$ ne change pas vraiment $Y_n$ (c'est juste un foncteur d'oubli). En particulier, pour tout $m \geq n$,

$$P_m\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) = P_n\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \ .$$

En imaginant qu'un $P_\infty$ limite des $P_m$ puisse être construit, on se débarasse du $P_n$ dans le théorème en envoyant $m$ à l'infini.

# PARTIE IV
## LE CAS DISCRET INFINI

Le cours sur les séries ayant lieu en parallèle du cours de probabilité il est difficile de parler du cas discret infini en début de cours. Mais en fin de cours on peut essayer au moins de présenter la loi géométrique et la loi de Poisson et la loi forte des grands nombres. Par discret infini on entend dénombrable infini.

### 37. Généralités

On commence par la définition des dénombrables infinis et des probabilités sur ce type d'espace.

**Définition 37.1.** *Un ensemble $\Omega$ est dit dénombrable infini s'il est en bijection avec un sous ensemble infini de $\mathbb{N}$, et donc s'il est possible d'écrire $\Omega$ sous la forme d'une suite infinie $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$. On appelle probabilité sur $(\Omega, \mathcal{P}(\Omega))$ toute application $P : \mathcal{P}(\Omega) \to [0,1]$ qui vérifie les deux points suivants:*

*(1) $P(\Omega) = 1$,*

*(2) $\forall A_i \in \mathcal{P}(\Omega)$, avec $i \in I$ et $I \subset \mathbb{N}^\star$, si $A_i \cap A_j = \emptyset$ pour tous $i \neq j$ dans $I$, alors*

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i) \ .$$

*Le triplet $(\Omega, \mathcal{P}(\Omega), P)$ est appelé espace de probabilité dénombrable infini. On dit que $\Omega$ est l'univers des possibles (ou encore l'ensemble fondamental). Les événements sont les éléments $A \in \mathcal{P}(\Omega)$. Le réel $P(A)$ est appelé probabilité de l'événement $A$.*

Les formules de la Partie II restent toutes valables dès lors que l'on comprend qu'on parle de séries et non plus de sommes finies. La donnée d'une probabilité $P : \mathcal{P}(\Omega) \to [0,1]$ est ainsi équivalente à la donnée d'une suite de réels $(p_n)_n \in [0,1]$, vérifiant que $\sum_{n=1}^{+\infty} p_n = 1$, via la formule $P(\{\omega_i\}) = p_i$ pour tout $i = 1, \dots, n$, si $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$. On a alors

$$P(A) = \sum_{i\in X_A} p_i \tag{37.1}$$

pour tout événement $A \in \mathcal{P}(\Omega)$, où, dans la formule ci-dessus, $X_A = \{i \ / \ \omega_i \in A\}$. En d'autres termes, une probabilité est entièrement caractérisée par la connaissance des $p_i = P(\{\omega_i\})$, qui vérifient forcément que $\sum p_n$ est convergente et vaut 1. La somme dans (37.1) est une somme de série si $A$ est infini.

### 38. La loi géométrique et la loi de Poisson

Deux lois classiques se rajoutent à celles déjà vues: la loi géométrique et la loi de Poisson. Elles ont toutes deux pour cadre les probabilités discrètes infinies.

**Définition 38.1.** *Soit $p \in ]0,1]$. On appelle loi géométrique de paramètre $p$ la probabilité sur $(\mathbb{N}^\star, \mathcal{P}(\mathbb{N}^\star))$ donnée par*

$$P(\{k\}) = p(1-p)^{k-1}$$

*pour tout $k \in \mathbb{N}^\star$.*

Si $p = 1$ on est dans le cas dégénéré où $P(\{k\}) = 0$ pour tout $k \geq 2$ et $P(\{1\}) = 1$. Pour vérifier qu'il s'agit bien d'une loi de probabilité, il faut vérifier que

$$p \sum_{n=1}^{+\infty} (1-p)^{n-1} = 1 \ . \tag{38.1}$$

Pour $a \in \mathbb{R}\backslash\{1\}$, on a la formule

$$\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a} \tag{38.2}$$

pour tout $n \geq 1$. Celle-ci se démontre facilement par récurrence sur $n$. Par suite,

$$\begin{aligned} \sum_{n=1}^{N} (1-p)^{n-1} &= \sum_{n=0}^{N-1} (1-p)^n \\ &= \frac{1}{p}\left(1-(1-p)^N\right) \ . \end{aligned}$$

Comme $0 < p \leq 1$ on a $0 \leq 1-p < 1$ et donc $(1-p)^N \to 0$ lorsque $N \to +\infty$. On en déduit que la série $\sum_{n\geq 1}(1-p)^{n-1}$ converge et que sa somme vaut $\frac{1}{p}$. D'où la relation (38.1).

Dans le cas dénombrable infini l'existence de l'espérance et de la variance ne sont plus garanties. Il faut des conditions de convergence des séries $\sum kP(\{k\})$ pour l'espérance et $\sum k^2 P(\{k\})$ pour la variance (existence des moments d'ordre 1 et 2). Ces conditions sont remplies dans le cas de la loi géométrique. Le calcul explicit de l'espérance et de la variance dans ce cadre fait appel à la théorie des séries entières.

**Proposition 38.1** (Espérance et variance pour la loi géométrique)**.** *Si une variable aléatoire $X : \Omega \to \mathbb{N}^\star$ suit la loi géométrique de paramètre $p$, alors $\mathbb{E}(X) = \frac{1}{p}$ et $Var(X) = \frac{1-p}{p^2}$.*

*Démonstration.* On définit l'espérance et la variance comme on l'a fait dans la Partie III, à ceci près que l'on parle maintenant de séries (et qu'il faut donc s'assurer de leur convergence). L'espérance est donnée par

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} nP(X = n)$$

et pour la variance on utilise la formule de Huygens $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Posons $\overline{p} = 1-p$ de sorte que $0 \leq \overline{p} < 1$. La série entière $\sum x^n$ a pour rayon de convergence $R = 1$ et, en vertue de (38.2), sa somme vaut

$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x} \ .$$

On peut dériver terme à terme dans $]-1, 1[$, et donc en particulier en $\overline{p}$. Pour $-1 < x < 1$,

$$\begin{aligned} \frac{1}{(1-x)^2} &= \frac{d}{dx}\left(\frac{1}{1-x}\right) \\ &= \sum_{n=1}^{+\infty} nx^{n-1} \ . \end{aligned} \tag{38.3}$$

Donc, en particulier,

$$\mathbb{E}(X) = p\sum_{n=1}^{+\infty} n\overline{p}^{n-1} = \frac{p}{(1-\overline{p})^2} \ .$$

On en déduit que $\mathbb{E}(X) = \frac{1}{p}$. Pour calculer $\mathbb{E}(X^2)$ on dérive essentiellement deux fois la série entière. Il suit de (38.3) que

$$\frac{x}{(1-x)^2} = \sum_{n=1}^{+\infty} nx^n \ .$$

On a là encore une série entière de rayon de convergence 1. On dérive de nouveau dans $]-1,1[$. On obtient que

$$\begin{aligned} \frac{x+1}{(1-x)^3} &= \frac{d}{dx}\left(\frac{x}{(1-x)^2}\right) \\ &= \sum_{n=1}^{+\infty} n^2 x^{n-1} \ . \end{aligned} \tag{38.4}$$

et on trouve ainsi avec (38.4) que

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{n=1}^{+\infty} n^2 P(X=n) \\ &= p\sum_{n=1}^{+\infty} n^2 \overline{p}^{n-1} \\ &= \frac{p(\overline{p}+1)}{(1-\overline{p})^3} \\ &= \frac{2-p}{p^2} \ . \end{aligned}$$

Avec la formule de Huygens on obtient alors que

$$\mathrm{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} \ .$$

Le théorème est démontré. □

La loi géométrique modélise une situation où l'on attend le premier succès (on parle de temps d'attente). Supposons que l'on ait une urne qui contient 6 boules blanches, 4 boules rouges et 2 boules noires. On effectue des tirages successifs avec remise jusqu'à obtenir une boule rouge. La variable aléatoire $X$ considérée donne le nombre de tirages effectués. Notons $B$ pour boule blanche, $R$ pour boule rouge et $N$ pour boule noire. Ici $\Omega = \{B, R, N\}^{\mathbb{N}^\star}$, qui est dénombrable infini, et $X : \Omega \to \mathbb{N}^\star \cup \{+\infty\}$ est donné par: $X(\omega) =$ numéro du premier rang auquel on voit le $R$ apparaître, étant entendu que $X(\omega) = +\infty$ si $\omega$ ne comporte aucun $R$. Formellement l'ensemble d'arrivée n'est pas vraiment $\mathbb{N}^\star$ mais on peut quand même décrire les $\{X = k\}$ pour $k \in \mathbb{N}^\star$. Il y a équiprobabilité à chaque tirage. L'événement $\{X = k\}$ signifie que les $k-1$ premières boules tirées sont dans $\{B, N\}$ tandis que la $k$ème boule tirée est $R$. Les tirages sont indépendants. La probabilité de tirer un $R$ est $p = \frac{4}{12} = \frac{1}{3}$. La probabilité de ne pas tirer un $R$ est $1-p$. Donc $P(X = k) = p(1-p)^{k-1}$ et on retrouve bien l'expression de la loi géométrique de paramètre $p = \frac{1}{3}$.

La probabilité $P(X = k)$ correspond à la probabilité d'obtenir, dans une succession d'épreuves de Bernoulli indépendantes, $k - 1$ échecs suivi d'une succès. C'est une loi sans mémoire au sens où pour tous $m, n \in \mathbb{N}$,

$$P(X > m + n/X > m) = P(X > n) \tag{38.5}$$

qui est la définition de la perte de mémoire en probabilité. Pourquoi sans mémoire ? Imaginons que $X$ modélise la durée de vie. Alors (38.5) exprime que la probabilité de vivre au moins encore $n$ années sachant qu'on en a déjà vécues au moins $m$ ne dépend pas de $m$. On a

$$P(X > m + n/X > m) = \frac{P(X > m + n)}{P(X > m)} \tag{38.6}$$

car $\{X > m + n\} \cap \{X > m\} = \{X > m + n\}$. On a par ailleurs, en vertue de (38.1), que

$$\begin{aligned} P(X > k) &= \sum_{n=k+1}^{+\infty} P(X = n) \\ &= p(1-p)^k \sum_{n=1}^{+\infty} (1-p)^{n-1} \\ &= (1-p)^k \end{aligned}$$

Donc $P(X > n) = (1 - p)^n$, $P(X > m) = (1 - p)^m$, $P(X > m + n) = (1 - p)^{m+n}$ et on obtient bien avec (38.6) que (38.5) est vraie. D'où la perte de mémoire de la loi géométrique.

**Définition 38.2.** *Soit $\theta > 0$. La loi de Poisson $\mathcal{P}(\theta)$ de paramètre $\theta$ est la probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ donnée par la loi $P(\{k\}) = \frac{\theta^k}{k!} e^{-\theta}$ pour tout $k \in \mathbb{N}$.*

Le développement en série entière de l'exponentielle assure qu'il sagit bien là d'une loi de probabilité au sens où $\sum_{k \geq 0} P(\{k\}) = 1$. Si $X$ est une variable aléatoire qui suit la loi $\mathcal{P}(\theta)$ on calcule facilement $\mathbb{E}(X) = \theta$ et $\mathrm{Var}(X) = \theta$ puisque, par exemple,

$$\sum_{n=0}^{+\infty} nP(X = n) = e^{-\theta} \sum_{n=1}^{+\infty} \frac{\theta^n}{(n-1)!} = \theta e^{-\theta} \sum_{n=0}^{+\infty} \frac{\theta^n}{n!} = \theta \ .$$

La loi de Poisson est d'importance car elle peut être vue comme une approximation de la loi binomiale $\mathcal{B}(n, p)$ lorsque $n, p \to +\infty$ et sont liés par une relation $np = \theta$.

**Théorème 38.1.** *Soit $\theta > 0$ et soit $n \in \mathbb{N}^\star$. On pose $p_n = \frac{\theta}{n}$. Alors la loi binomiale $\mathcal{B}(n, p_n)$ converge vers la loi de Poisson $\mathcal{P}(\theta)$ lorsque $n \to +\infty$ au sens où pour tout $k \in \mathbb{N}$, $P_n(\{k\}) \to P(\{k\})$ si $P_n$ est la probabilité de la loi binomiale $\mathcal{B}(n, p_n)$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, et $P$ la probabilité de la loi de Poisson.*

*Démonstration.* On écrit que

$$\begin{aligned} P_n(\{k\}) &= \frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} \\ &= (1 - p_n)^n \times \frac{n^k p_n^k}{k!} \times \frac{n!}{n^k (n-k)!} \times (1 - p_n)^{-k} \ . \end{aligned}$$

Par développement limité, $(1-p_n)^n \to e^{-\theta}$ puisque $np_n = \theta$. On a aussi que $(1-p_n)^{-k} \to 1$ tandis que $\frac{n^k p_n^k}{k!} = \frac{\theta^k}{k!}$. La formule de Stirling donne que

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \ .$$

On en déduit que

$$\begin{aligned}\frac{n!}{n^k(n-k)!} &\sim e^{-k}\left(\frac{n}{n-k}\right)^{n-k} \\ &\sim e^{-k}\left(1+\frac{k}{n-k}\right)^{n-k}\end{aligned}$$

et, de nouveau par développement limité, on obtient que

$$\lim_{n\to+\infty} \frac{n!}{n^k(n-k)!} = 1 \ .$$

D'où le résultat. □

On calcule pour finir les espérances et variances des variables aléatoires qui suivent une loi de Poisson.

**Proposition 38.2** (Espérance et variance pour la loi de Poisson)**.** *Si une variable aléatoire $X : \Omega \to \mathbb{N}$ suit la loi de Poisson de paramètre $\theta$, alors $\mathbb{E}(X) = \theta$ et $Var(X) = \theta$.*

*Démonstration.* On calcule

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{+\infty} kP(\{k\}) \\ &= \theta e^{-\theta} \sum_{k=1}^{+\infty} \frac{\theta^{k-1}}{(k-1)!} \\ &= \theta e^{-\theta} \sum_{k=0}^{+\infty} \frac{\theta^k}{k!}\end{aligned}$$

et donc $\mathbb{E}(X) = \theta$. On a aussi

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=1}^{+\infty} k^2 P(\{k\}) \\ &= e^{-\theta} \sum_{k=1}^{+\infty} \frac{k\theta^k}{(k-1)!} \\ &= \theta e^{-\theta} \sum_{k=0}^{+\infty} \frac{(1+k)\theta^k}{k!} \\ &= \theta e^{-\theta} \left(\sum_{k=0}^{+\infty} \frac{\theta^k}{k!} + \sum_{k=1}^{+\infty} \frac{\theta^k}{(k-1)!}\right) \\ &= \theta e^{-\theta} \left(e^{\theta} + \theta e^{\theta}\right)\end{aligned}$$

et donc $\mathbb{E}(X^2) = \theta(1+\theta)$. Par suite, avec la formule de Huygens, on obtient que $\mathrm{Var}(X) = \theta(1+\theta) - \theta^2 = \theta$. La proposition est démontrée. □

# PARTIE V
## LOIS DES GRANDS NOMBRES

Il s'agit d'une partie introductive au cours que vous verrez les années suivantes où vous étudierez beaucoup plus rigoureusement les lois faibles et fortes des grands nombres. Le cadre continu, beaucoup plus riche, est le bon cadre pour cela. Lorsque $\Omega$ se situe dans le cadre continu, typiquement $\Omega = \mathbb{R}$, on ne peut pas prendre $\mathcal{P}(\Omega)$ comme espace de tous les événements, la raison étant que tout sous ensemble de $\mathbb{R}$ n'est pas forcément mesurable. Un sens précis à cette affirmation sera donné en L3. Il faut restreindre la classe des événements et pour cela introduire la notion de tribu. Une partie $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu si

(1) $\Omega \in \mathcal{A}$,

(2) $\forall A \in \mathcal{A}$ on a $A^c \in \mathcal{A}$,

(3) $\forall (A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$, on a $\displaystyle\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Lorsque $\Omega = \mathbb{R}$, on montre qu'il existe une plus petite tribu $\mathcal{B}$ contenant les intervalles de $\mathbb{R}$. Cette tribu est appelée tribu de Borel de $\mathbb{R}$ (ou tribu borélienne de $\mathbb{R}$). Une probabilité sur $(\Omega, \mathcal{A})$, où $\mathcal{A}$ est une tribu sur $\Omega$, est alors une fonction $P : \mathcal{A} \to [0, 1]$ vérifiant que

(1) $P(\Omega) = 1$,

(2) $\forall (A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}} \ / \ \forall i \neq j, A_i \cap A_j = \emptyset,$ on a $P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n)$ .

Les éléments de $\mathcal{A}$ sont les événements et on va alors retrouver un certain nombre des compagnons vus dans les sections précédentes. On montre que nécessairement $P(\emptyset) = 0$, que (2) reste vraie si on ne considère qu'un nombre fini d'événements incompatibles, que

$$P(A) + P(A^c) = 1$$

pour tout $A \in \mathcal{A}$, que $P(A_1) \leq P(A_2)$ si $A_1 \subset A_2$), et que

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

pour tous $A_1, A_2 \in \mathcal{A}$. Les lois de Morgan permettant d'écrire que $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c$, on a bien que $A_1 \cap A_2 \in \mathcal{A}$. On montre aussi facilement que si $(A_n)$ est une suite croissante (resp. décroissante) d'événements, alors

$$P(\lim_{n\to+\infty} A_n) = \lim_{n\to+\infty} P(A_n) \ . \tag{38.7}$$

Le cas décroissant se déduit du cas croissant par passage aux complémentaires. Pour démontrer la proposition dans le cas croissant on pose $B_0 = A_0$ puis $B_n = A_n \cap A_{n-1}^c$ (couronnes). Ces événements sont deux à deux disjoints et tels que $A_n = \bigcup_{m=0}^n B_m$. Donc

$$P(A_n) = \sum_{m=0}^{n} P(B_m) \ .$$

On montre sans difficulté que

$$\lim_{n\to+\infty} A_n = \bigcup_{n\geq 0} A_n = \bigcup_{n\geq 0} B_n$$

et donc

$$P(\lim_{n\to+\infty} A_n) = \sum_{n\geq 0} P(B_n)$$

puisque les $B_n$ sont deux à deux disjoints. D'où

$$P(\lim_{n\to+\infty} A_n) = \lim_{n\to+\infty} \sum_{m=0}^{n} P(B_m)$$

et comme $\sum_{m=0}^{n} P(B_m) = P(A_n)$ on a démontré le résultat voulu, à savoir (38.7).

## 39. Variables aléatoires à densité

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Une variable aléatoire réelle sur $(\Omega, \mathcal{A}, P)$ est une application $X : \Omega \to \mathbb{R}$ qui vérifie que

$$\forall I \text{ intervalle de } \mathbb{R},\ \{X \in I\} \in \mathcal{A}\ .$$

La loi de $X$ est alors la probabilité $P_X$ sur $\mathbb{R}$ donnée par $P_X(B) = P(X \in B)$ pour tout borélien $B$ de $\mathbb{R}$. Elle est caractérisée par la fonction de répartition

$$F_X(x) = P(X \leq x)\ .$$

Une variable aléatoire $X$ est dite à densité $p : \mathbb{R} \to \mathbb{R}^+$ si $p$ est relativement propre pour qu'on puisse l'intégrer (ce qui sera rendu plus clair en L3), par exemple continue par morceaux ne présentant qu'un nombre fini de discontinuités, si elle est intégrable sur $\mathbb{R}$ et telle que

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

(vous verrez ce que cela signifie dans le cours d'intégration de L2 puis dans le cours de L3), et si elle est telle que

$$F_X(x) = \int_{-\infty}^{x} p(t)dt\ .$$

Si $I$ est un intervalle d'extrémités $A < B$, les relations ci-dessus et les relations de Chasles pour l'intégrale donnent que

$$P(X \in I) = \int_{A}^{B} p(t)dt\ .$$

Un exemple de densité possible est la densité de la loi uniforme sur $[a, b]$, avec $a < b$, qui est donnée par

$$p(x) = \frac{1}{b-a}\mathbb{1}_{]a,b[}\ .$$

Un autre exemple est la densité de la loi gaussienne

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ .$$

Une variable alétaoire réelle $X$ à densité $p$ est dite avoir un moment d'ordre $m \in \mathbb{N}^\star$ si

$$\int_{-\infty}^{+\infty} |t|^m p(t)dt < +\infty\ .$$

Le moment d'ordre $m$ est alors noté

$$\mathbb{E}(X^m) = \int_{-\infty}^{+\infty} |t|^m p(t)dt \ .$$

Lorsque $X$ est à densité et a un moment d'ordre 1 on dit qu'elle est intégrable et on peut définir l'espérance de $X$ par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} |t| p(t)dt \ .$$

On peut aussi montrer (car $\int p = 1$) que si $X$ a un moment d'ordre $n$ alors elle a des moments d'ordre $m$ pour tout $m \leq n$. En particulier, si $X$ a un moment d'ordre 2, on dit qu'elle est de carré intégrable, alors elle est aussi intégrable et on peut définir la variance

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

(formule de Huygens). L'espérance est linéaire et la variance se transforme de façon connue par transport affine:

(1) (Linéarité de l'espérance) Si $X_1, \dots, X_N$ sont des variables aléatoires réelles intégrables

$$\mathbb{E}(\sum_{i=1}^N \alpha_i X_i) = \sum_{i=1}^N \alpha_i \mathbb{E}(X_i) \tag{39.1}$$

pour tous $\alpha_1, \dots, \alpha_N \in \mathbb{R}$.

(2) (Transport affine de la variance) Si $X$ est de carré intégrable,

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X) \tag{39.2}$$

pour tous $a, b \in \mathbb{R}$.

Les inégalités de Markov et de Bienaymé-Tchebychef du cas discret restent vraies:

(1) Inégalité de Bienaymé-Tchebychef: Si $X$ est une variable aléatoire réelle de carré intégrable, alors

$$P\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{1}{a^2} \mathrm{Var}(X) \tag{39.3}$$

pour tout $a > 0$.

(2) Inégalité de Markov: Si $X$ est une variable aléatoire réelle intégrable à valeurs positives ou nulles, alors

$$P(X \geq x) \leq \frac{1}{x} \mathbb{E}(X) \tag{39.4}$$

pour tout $x > 0$.

Reste à définir l'indépendance de familles de variables aléatoires dans le cas continu. Deux variables aléatoires réelles $X, Y$ sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ sont dites indépendantes si pour tous intervalles $I, J \subset \mathbb{R}$,

$$P(\{X \in I\} \cap \{Y \in J\}) = P(X \in I) P(Y \in J) \ .$$

Considérons maintenant $X_1, \dots, X_N$ des variables aléatoires. Les $X_i$ sont dites deux à deux indépendantes si $X_i$ et $X_j$ sont indépendantes pour tous $i \neq j$. Elles sont dites mutuellement indépendantes si pour tous intervalles $I_1, \dots, I_N$ de $\mathbb{R}$,

$$P\left(\bigcap_{i=1}^N \{X_i \in I_i\}\right) = \prod_{i=1}^N P(\{X_i \in I_i\}) \ .$$

Si on parle d'une suite $(X_n)_{n\in\mathbb{N}^\star}$, les $X_n$ sont dites mutuellement indépendantes si pour tout $N \in \mathbb{N}^\star$ les $X_1, \dots, X_N$ sont mutuellement indépendantes. Des variables aléatoires mutuellement indépendantes sont deux à deux indépendantes, mais la réciproque est fausse. On montre que:

(1) Si $X_1, \dots, X_N$ sont des variables aléatoires réelles à densité mutuellement indépendantes, et si les $n_1, \dots, n_N \in \mathbb{N}^\star$ son tels que $X_i$ a un moment d'ordre $n_i$ pour tout $i = 1, \dots, N$, alors

$$\mathbb{E}(X_1^{n_1} \dots X_N^{n_N}) = \prod_{i=1}^{N} \mathbb{E}(X_i^{n_i}) \ . \tag{39.5}$$

(2) Si $X_1, \dots, X_N$ sont des variables aléatoires réelles à densité de carré intégrable et deux à deux indépendantes, alors

$$\mathrm{Var}(X_1 + \cdots + X_N) = \sum_{i=1}^{N} \mathrm{Var}(X_i) \ . \tag{39.6}$$

Les preuves de tous ces résultats sont essentiellement les mêmes que dans le cas discret.

## 40. Le problème du cas discret

On illustre ici l'idée qu'il est impossible de trouver une suite de variables aléatoires réelles deux à deux indépendantes dans le cas fini, hors cas triviaux, ce qui explique pourquoi la loi faible des grands nombres, telle qu'énoncée à la section suivante, pose problème dans ce cadre. On dira d'un espace de probablitié fini $(\Omega, \mathcal{P}(\Omega), P)$ qu'il est non troué si $P(\{\omega\}) > 0$ pour tout $\omega \in \Omega$.

**Lemme 40.1.** *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini non troué. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles deux à deux indépendantes sur $(\Omega, \mathcal{P}(\Omega), P)$. Alors, sauf pour peut-être un nombre fini d'entre elles, les $X_n$ sont toutes constantes.*

*Démonstration.* Il suffit de montrer qu'il n'existe pas de suite $(X_n)_{n\geq 1}$ de variables aléatoires réelles deux à deux indépendantes sur $(\Omega, \mathcal{P}(\Omega), P)$ qui soit telle que $X_n$ prenne au moins deux valeurs pour tout $n$. Par l'absurde, on suppose qu'une telle suite existe. On note $X_n : \Omega \to E_n$. En vertu du Théorème 34.1, pour toute suite $(f_n)_{n\geq 1}$ de fonctions, $f_n : E_n \to \mathbb{R}$, la suite $(f_n(X_n))_{n\geq 1}$ est encore une suite de variables aléatoires réelles deux à deux indépendantes. Les $X_n$ prenant toutes au moins deux valeurs, il est facile de trouver des $f_n$ telle que $f_n(X_n) : \Omega \to \{0, 1\}$ soit surjective (i.e prenne les deux valeurs 0 et 1) pour tout $n$. Soit $k = \sharp\Omega$ et soit $\tilde{X}_n = f_n(X_n)$. Il y a $2^k$ applications possibles de $\Omega$ dans $\{0, 1\}$. Donc, forcément, il existe $n_1 \neq n_2$ tels que $\tilde{X}_{n_1} = \tilde{X}_{n_2}$. Notons

$$p_0 = P(\{\tilde{X}_{n_1} = 0\}) = P(\{\tilde{X}_{n_2} = 0\}) \ , \ p_1 = P(\{\tilde{X}_{n_1} = 1\}) = P(\{\tilde{X}_{n_2} = 1\}) \ .$$

Comme $\{\tilde{X}_{n_1} = 0\} \cap \{\tilde{X}_{n_2} = 0\} = \{\tilde{X}_{n_1} = 0\}$, et même chose avec 1, l'indépendance de $\tilde{X}_{n_1}$ et $\tilde{X}_{n_2}$ donne que $p_0 = p_0^2$ et $p_1 = p_1^2$. On a aussi $p_0 + p_1 = 1$. Donc soit $p_0 = 0$ et $p_1 = 1$, soit $p_0 = 1$ et $p_1 = 0$. Supposons par exemple que $p_0 = 0$ et $p_1 = 1$. Comme $(\Omega, \mathcal{P}(\Omega), P)$ est non troué, $\{\tilde{X}_{n_1} = 0\} = \emptyset$ et ainsi $\tilde{X}_{n_1}$ ne prend que la seule valeur 1, une contradiction avec notre construction. D'où le lemme. $\square$

## 41. Lois des grands nombres

On peut maintenant aborder les deux formes faibles et fortes de la loi des grands nombres. La loi faible est basée sur l'inégalité de Bienaymé-Tchebychef. La loi forte sur l'inégalité de Markov.

**Théorème 41.1** (Loi faible des grands nombres)**.** *Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles à densité deux à deux indépendantes, de même loi (même densité) et de carré intégrable. Pour tout $n \geq 1$ on pose*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ . \tag{41.1}$$

*Alors la suite des $\overline{X}_n$ converge en probabilité vers $\mathbb{E}(X_1)$ au sens où*

$$\lim_{n\to+\infty} P\left(\left|\overline{X}_n - \mathbb{E}(X_1)\right| \geq \varepsilon\right) = 0 \tag{41.2}$$

*pour tout $\varepsilon > 0$. La moyenne $\overline{X}_n$ est appelée moyenne empirique des $X_1, \ldots, X_n$. La moyenne empirique $\overline{X}_n$ converge donc en probabilité vers $\mathbb{E}(X_1)$ dès lors que les $X_n$ sont deux à deux indépendantes, de carré intégrable et de même loi.*

*Démonstration.* Ce théorème très parlant n'est rien d'autre qu'une conséquence dirècte de l'inégalité de Bienaymé-Tchebychef et des propriétés de l'espérance. Par linéarité de l'espérance (39.1), et puisque les $X_n$ suivent toutes la même loi, $\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1)$ pour tout $n$. Soit $\varepsilon > 0$ fixé. D'après l'inégalité de Bienaymé-Tchebychef (39.3) on a alors que

$$P\left(\left|\overline{X}_n - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathrm{Var}(\overline{X}_n) \tag{41.3}$$

pour tout $n$. En vertue de (39.2) et (39.6) on a que

$$\begin{aligned} \mathrm{Var}(\overline{X}_n) &= \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}\mathrm{Var}(X_1) \end{aligned} \tag{41.4}$$

puisque, les $X_n$ suivant tous la même loi, on a aussi que $\mathrm{Var}(X_n) = \mathrm{Var}(X_1)$ pour tout $n$. En combinant (41.3) et (41.4) on obtient (41.2). La loi faible des grands nombres est démontrée. □

La loi forte des grands nombres donne une convergence plus forte.

**Théorème 41.2** (Loi forte des grands nombres)**.** *Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles à densité mutuellement indépendantes sur $(\Omega, \mathcal{A}, P)$. On suppose que les $X_n$ suivent toutes la même loi (même densité) et qu'elles sont intégrables. Soit $\overline{X}_n$ donnée par* (41.1)*. Alors la suite des $\overline{X}_n$ converge presque sûrement vers $\mathbb{E}(X_1)$ au sens où si $\mathbb{L}$ désigne l'événement constitué des $\omega \in \Omega$ pour lesquels la limite $\lim\limits_{n\to+\infty} \overline{X}_n(\omega)$ existe et vaut $\mathbb{E}(X_1)$, alors $P(\mathbb{L}) = 1$.*

La preuve de ce théorème est compliquée et vous la verrez en toute généralité en L3. Il existe néanmoins une preuve simple si l'on suppose que les variables aléatoires ont des moments d'ordre 4.

*Démonstration.* Quitte à changer $X_n$ par $X_n - \mathbb{E}(X_1)$ on peut supposer que les variables aléatoires $X_n$ sont centrées, à savoir que $\mathbb{E}(X_n) = 0$ pour tout $n$, et il s'agit maintenant de montrer que les moyennes empiriques $\overline{X}_n$ convergent presque sûrement vers 0. Soit

$$\mathbb{L} = \left\{ \omega \in \Omega \ / \ \lim_{n \to +\infty} \overline{X}_n(\omega) \text{ existe et vaut } 0 \right\} .$$

On suppose ici que les variables aléatoires ont un moment d'ordre 4, et donc que $\mathbb{E}(X_1^4) < +\infty$. Cela simplifie grandement la démonstration. On va travailler avec la quantité $\overline{X}_n^4$. On a

$$\mathbb{E}(\overline{X}_n^4) = \frac{1}{n^4} \sum_{i_1, i_2, i_3, i_4 = 1}^{n} \mathbb{E}(X_{i_1} X_{i_2} X_{i_3} X_{i_4})$$

et par indépendance des $X_n$, comme $\mathbb{E}(X_i) = 0$ pour tout $i$, on obtient avec (39.5) que

$$\mathbb{E}(\overline{X}_n^4) = \frac{1}{n^4} \sum_{\substack{i_1, i_2 = 1 \\ i_1 \neq i_2}}^{n} \mathbb{E}(X_{i_1})^2 \mathbb{E}(X_{i_2})^2 + \frac{1}{n^4} \sum_{i=1}^{n} \mathbb{E}(X_i)^4 .$$

Maintenant, comme les $X_i$ suivent tous la même loi, on trouve que

$$\mathbb{E}(\overline{X}_n^4) = \frac{n-1}{n^3} \mathbb{E}(X_1^2)^2 + \frac{1}{n^3} \mathbb{E}(X_1^4)$$

Les $\mathbb{E}(\overline{X}_n^4)$ sont tous positifs ou nuls. On en déduit par comparaison avec la série de Riemann de terme général $\frac{1}{n^2}$ que $\sum \mathbb{E}(\overline{X}_n^4)$ est une série convergente. L'inégalité de Markov (39.4) donne alors que pour tout $n \geq 1$ et tout $M \geq 1$,

$$P\left( \sum_{i=1}^{n} \overline{X}_i^4 \geq M \right) \leq \frac{1}{M} \mathbb{E}\left( \sum_{i=1}^{n} \overline{X}_i^4 \right)$$

et comme

$$\mathbb{E}\left( \sum_{i=1}^{n} \overline{X}_i^4 \right) = \sum_{i=1}^{n} \mathbb{E}\left( \overline{X}_i^4 \right) \leq C$$

puisque la série $\sum \mathbb{E}(\overline{X}_n^4)$ converge, on obtient que

$$P\left( \sum_{i=1}^{n} \overline{X}_i^4 \geq M \right) \leq \frac{C}{M} \tag{41.5}$$

pour tous $n \geq 1$ et $M \geq 1$, où $C > 0$ ne dépend ni de $n$ ni de $M$. Notons $S = \sum_{n=1}^{+\infty} \overline{X}_n^4$ (possiblement $S(\omega) = +\infty$ pour certains $\omega$, mais d'après (41.5) il y a aussi des $\omega$ pour lesquel $S(\omega) < +\infty$). On a

$$\{S \leq M\} = \bigcap_{n \in \mathbb{N}^\star} \left\{ \sum_{i=1}^{n} \overline{X}_i^4 \leq M \right\} .$$

On a aussi $\left\{ \sum_{i=1}^{n_2} \overline{X}_i^4 \leq M \right\} \subset \left\{ \sum_{i=1}^{n_1} \overline{X}_i^4 \leq M \right\}$ si $n_2 \geq n_1$, et donc ces événements forment une suite décroissante en $n$ d'événements pour l'inclusion. Par suite, avec

(38.7),

$$\begin{aligned} P(S \leq M) &= \lim_{n \to +\infty} P\left(\left\{\sum_{i=1}^{n} \overline{X}_i^4 \leq M\right\}\right) \\ &= 1 - \lim_{n \to +\infty} P\left(\left\{\sum_{i=1}^{n} \overline{X}_i^4 > M\right\}\right) \\ &\geq 1 - \frac{C}{M} \end{aligned} \tag{41.6}$$

en vertue de (41.5). On a aussi que $\{S < +\infty\} = \bigcup_{M \in \mathbb{N}^*} \{S \leq M\}$ et les $\{S \leq M\}$ forment une suite croissante en $M$ d'événements pour l'inclusion. Par passage aux probabilités et par passage à la limite en $M \to +\infty$ on obtient avec (41.6) que $P(S < +\infty) = 1$ (car supérieur ou égale à 1). En d'autres termes, si $A$ est l'événement $A = \{S < +\infty\}$ constitué des $\omega$ pour lesquels la série de terme général $\overline{X}_n^4(\omega)$ converge, alors $P(A) = 1$. Or si $\sum \overline{X}_n^4(\omega)$ converge, alors forcément $\overline{X}_n^4(\omega) \to 0$ lorsque $n \to +\infty$ (le terme général d'une série convergente tend vers 0). Donc $A \subset \mathbb{L}$. Et comme $P(A) = 1$ c'est que $P(\mathbb{L}) = 1$. Le théorème est démontré. □

Ces théorèmes sont très élégants et faciles à démontrer car on a tout mis dans $\Omega$ (d'où la nécessité de passer en continu pour avoir suffisamment de place). Par contre on ne voit plus très bien en quoi ils modélisent une succession d'événements qui nécessite de passer aux espaces produits $\Omega$, $\Omega^2$, $\Omega^3$ etc. Pour cela il faut revenir à la discussion de la Section 36, ou alors se reporter au cours que vous suivrez en L3.

# PARTIE VI

## QUELQUES EXERCICES EN PLUS

Ce chapitre contient quelques exercices non triviaux et parfois surprenants.

**EXERCICE 30**:** La loi de Hardy-Weinberg, proposée en 1908 indépendamment par le mathématicien anglais Hardy et le médecin allemand Weinberg, décrit les relations entre les fréquences génotypiques et les fréquences alléliques. Considérons un gène qui se présente sous deux allèles $A$ et $a$. Un individu possède deux allèles d'un même gène, donc $AA$, $Aa$ ou $aa$. Par passage de génération, l'enfant reçoit un allèle de chacun de ses parents au hasard. On note $p$, $q$, $r$ les proportions de génotypes dans une génération et $P$, $Q$, $R$ les proportions dans la génération suivante. Montrer que $Q^2 = 4PR$. C'est la loi de Hardy-Weinberg: "les proportions des différents génotypes restent constantes d'une génération à l'autre".

CORRECTION: On utilise la formule généralisée des probabilités totales. On considère les événements

$A_1$ "le père est $AA$ et la mère est $AA$"
$A_2$ "le père est $AA$ et la mère est $Aa$"
$A_3$ "le père est $AA$ et la mère est $aa$"
$A_4$ "le père est $Aa$ et la mère est $AA$"
$A_5$ "le père est $Aa$ et la mère est $Aa$"
$A_6$ "le père est $Aa$ et la mère est $aa$"
$A_7$ "le père est $aa$ et la mère est $AA$"
$A_8$ "le père est $aa$ et la mère est $Aa$"
$A_9$ "le père est $aa$ et la mère est $aa$"

L'univers des possibles $\Omega$ est constitué des triplets "(génotype du père,génotype de la mère,génotype de l'enfant)". Les $A_i$ forment un système complet d'événements (ils agissent uniquement sur les deux premières variables du triplet, la troisième est laissée libre). On considère maintenant les événements $B$ "un individu de la deuxième génération est de génotype $AA$", $C$ "un individu de la deuxième génération est de génotype $Aa$" et $D$ "un individu de la deuxième génération est de génotype $aa$". Ces nouveaux événements agissent uniquement sur la troisième variable du couple, les deux premières sont laissées libres. Les deux premières variables sont gérées par les loi sd eprobabilités $p, q, r$ et on peut, sur ces deux premières variables, appliquer les probabilités conditionnelles. La troisième variable est gérée par l'équiprobabilité. On a

$$\begin{aligned}
P &= P(B) \\
&= \sum_{i=1}^{9} P(B/A_i)P(A_i) \\
&= P(A_1) + \frac{1}{2}P(A_2) + 0P(A_3) + \frac{1}{2}P(A_4) + \frac{1}{4}P(A_5) + 0\sum_{i=6}^{9} P(A_i) \\
&= p^2 + \frac{1}{2}pq + \frac{1}{2}pq + \frac{1}{4}q^2 \\
&= \left(p + \frac{q}{2}\right)^2 .
\end{aligned}$$

De même,

$$\begin{aligned}
Q &= P(C) \\
&= \sum_{i=1}^{9} P(C/A_i)P(A_i) \\
&= 0P(A_1) + \frac{1}{2}P(A_2) + P(A_3) + \frac{1}{2}P(A_4) + \frac{1}{2}P(A_5) \\
&\quad + \frac{1}{2}P(A_6) + P(A_7) + \frac{1}{2}P(A_8) + 0P(A_9) \\
&= \frac{1}{2}pq + pr + \frac{1}{2}pq + \frac{1}{2}q^2 + \frac{1}{2}qr + pr + \frac{1}{2}qr \\
&= pq + 2pr + qr + \frac{1}{2}q^2 \\
&= 2\left(p + \frac{q}{2}\right)\left(r + \frac{q}{2}\right) \ .
\end{aligned}$$

Enfin,

$$\begin{aligned}
R &= P(D) \\
&= \sum_{i=1}^{9} P(D/A_i)P(A_i) \\
&= 0\sum_{i=1}^{4} P(A_i) + \frac{1}{4}P(A_5) + \frac{1}{2}P(A_6) + 0P(A_7) + \frac{1}{2}P(A_8) + P(A_9) \\
&= \frac{1}{4}q^2 + \frac{1}{2}qr + \frac{1}{2}qr + r^2 \\
&= \left(r + \frac{q}{2}\right)^2 \ .
\end{aligned}$$

On a alors

$$\begin{aligned}
Q^2 &= 4\left(p + \frac{q}{2}\right)^2\left(r + \frac{q}{2}\right)^2 \\
&= 4PR
\end{aligned}$$

qui est la relation recherchée.

**EXERCICE 31⋆:** La famille Leclerc a 2 enfants. On considère les deux événements suivants. A savoir A "il y a deux enfants de sexes différents chez les Leclerc" et B "La famille Leclerc a au plus une fille". Les événements $A$ et $B$ sont-ils indépendants ? Qu'en serait-il si la famille Leclerc avait 3 enfants ? Et si la famille Leclerc avait $n$ enfants, $n \geq 4$ ?

CORRECTION: On considère que les probabilités d'avoir une fille ou un garçon sont les mêmes, et donc que l'on se situe dans une situation d'équiprobabilité. Si la famille Leclerc a $n$ enfants, $n = 2, 3, 4, \ldots$ l'univers des possibles est $\Omega = \{G, F\}^n$, où $G$ signifie garçon et $F$ signifie fille. Pour $n = 2$, $\sharp\Omega = 4$. L'événement $A$ est réalisé par les couples $(G, F)$ et $(F, G)$ et donc $P(A) = \frac{2}{4} = \frac{1}{2}$. Lorsque $n = 2$, $B$ est réalisé par les couples $(G, G)$, $(F, G)$ et $(G, F)$. Donc $P(B) = \frac{3}{4}$. Enfin, $A \cap B = A$ et donc $P(A \cap B) = \frac{1}{2}$. Donc $P(A \cap B) \neq P(A)P(B)$. Les événements $A$ et $B$ ne sont pas indépendants lorsque $n = 2$. Supposons $n = 3$. Alors $\sharp\Omega = 8$. On a $A = \Omega \backslash \{(F, F, F), (G, G, G)\}$ et donc $\sharp A = 6$. On

a $B = \{(F,G,G),(G,F,G),(G,G,F),(G,G,G)\}$ et donc $\sharp B = 4$. Enfin $A \cap B = \{(F,G,G),(G,F,G),(G,G,F)\}$ et donc $\sharp(A \cap B) = 3$. D'où $P(A \cap B) = \frac{3}{8}$, $P(A) = \frac{6}{8} = \frac{3}{4}$ et $P(B) = \frac{1}{2}$. On a bien $P(A \cap B) = P(A)P(B)$ et donc $A$ et $B$ sont indépendants lorsque $n = 3$. Supposons maintenant $n \geq 4$. On a $\sharp\Omega = 2^n$. On a toujours $A = \Omega \backslash \{(F,F,F),(G,G,G)\}$ et donc $\sharp A = 2^n - 2 = 2(2^{n-1} - 1)$. On a $B = \{(F,G,\ldots,G),(G,F,G,\ldots,G),\ldots,(G,\ldots,G,F),(G,\ldots,G)\}$ et donc $\sharp B = n+1$. On a enfin $A \cap B = B \backslash \{(G,\ldots,G)\}$ et donc $\sharp(A \cap B) = n$. Donc

$$\begin{aligned} P(A \cap B) = P(A)P(B) &\Leftrightarrow \frac{n}{2^n} = \left(1 - \frac{1}{2^{n-1}}\right)\frac{n+1}{2^n} \\ &\Leftrightarrow 2^{n-1}n = (2^{n-1} - 1)(n+1) \\ &\Leftrightarrow 2^{n-1} = n+1 \ . \end{aligned}$$

Soit $f$ la fonction $f(x) = 2^{x-1} - x - 1$. On a $f(4) = 3$ tandis que pour $x \geq 4$,

$$f'(x) = 2^{x-1}\ln(2) - 1 \geq 8\ln(2) - 1 > 0 \ .$$

Donc $f(x) \geq f(4)$ pour $x \geq 4$ et l'égalité ne peut avoir lieu. Ainsi, $A$ et $B$ ne sont de nouveau plus indépendants lorsque $n \geq 4$.

**EXERCICE 32:** On dispose de $n$ boîtes pouvant contenir de 0 à $n$ boules numérotées de 1 à $n$. On place les boules au hasard dans les $n$ boîtes. On note $p_n$ la probabilité que chaque boîte contienne exactement une boule. Que vaut $p_n$ ? Quelle est la limite de $p_n$ lorsque $n \to +\infty$ ?

CORRECTION: Ici $\Omega$ est l'ensemble des applications de $[\![1,n]\!]$ (les $n$ boules) dans $[\![1,n]\!]$ (les $n$ boites), et l'événement $A$ "chaque boîte contient exactement une boule" correspond aux bijections de $[\![1,n]\!]$ sur $[\![1,n]\!]$. On a donc $\sharp\Omega = n^n$ tandis que $\sharp A = n!$. Donc

$$p_n = \frac{n!}{n^n} \ .$$

La formule de Stirling donne que $n! \sim \sqrt{2\pi n}n^n e^{-n}$, et donc

$$p_n \sim \frac{\sqrt{2\pi n}}{e^n} \ .$$

L'exponentielle l'emporte sur toutes les puissances. On en déduit que $p_n \to 0$ lorsque $n \to +\infty$.

**EXERCICE 33:** Soient $n \in \mathbb{N}^\star$ et $a \in \mathbb{R}^{+\star}$. Soit $X$ une variable aléatoire dont la loi $P_X : [\![0,n]\!] \to [0,1]$ est donnée par

$$P(X = k) = \frac{a}{k+1}\binom{n}{k} \ .$$

Que vaut $a$ ? Calculer $\mathbb{E}(X+1)$. En déduire $\mathbb{E}(X)$.

CORRECTION: Il faut que $\sum_{k=0}^n P(X = k) = 1$. On veut donc que $a\sum_{k=0}^n \frac{1}{k+1}\binom{n}{k} = 1$. Or, cf. (6.2),

$$\frac{1}{k+1}\binom{n}{k} = \frac{1}{n+1}\binom{n+1}{k+1} \ .$$

Donc

$$\begin{aligned} a\sum_{k=0}^{n} P(X=k) &= \frac{a}{n+1}\sum_{k=0}^{n}\binom{n+1}{k+1} \\ &= \frac{a}{n+1}\sum_{k=1}^{n+1}\binom{n+1}{k} \\ &\frac{a}{n+1}\left(\sum_{k=0}^{n+1}\binom{n+1}{k}-1\right) \\ &= \frac{a}{n+1}\left((1+1)^{n+1}-1\right) \\ &= \frac{(2^{n+1}-1)a}{n+1}\ . \end{aligned}$$

On veut donc

$$a = \frac{n+1}{2^{n+1}-1}\ .$$

On a ensuite

$$\begin{aligned} \mathbb{E}(X+1) &= \sum_{k=0}^{n}(k+1)P(X=k) \\ &= a\sum_{k=0}^{n}\binom{n}{k} \\ &= a(1+1)^n = 2^n a\ . \end{aligned}$$

Or $\mathbb{E}(X+1) = \mathbb{E}(X)+1$ par linéarité de l'espérance. Ainsi

$$\mathbb{E}(X) = 2^n a - 1 = \frac{2^n(n-1)+1}{2^{n+1}-1}\ .$$

**EXERCICE 34:** Soit $n \in \mathbb{N}^\star$ et soit $X_n$ une variable aléatoire suivant la loi uniforme sur l'ensemble $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$. Que vaut la limite de $\mathbb{E}(X_n)$ lorsque $n \to +\infty$ ?

CORRECTION: On a par définition,

$$\mathbb{E}(X_n) = \sum_{k=1}^{n}\frac{k}{n}P(X_n = \frac{k}{n}) = \frac{1}{n^2}\sum_{k=1}^{n}k$$

puisque $P(X_n = \frac{k}{n}) = \frac{1}{n}$ par loi uniforme. Or $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$, et donc

$$\mathbb{E}(X_n) \to \frac{1}{2}$$

lorsque $n \to +\infty$.

**EXERCICE 35:** Soit $n \in \mathbb{N}^\star$ et soit $X_n$ une variable aléatoire suivant la loi binomiale de paramètres $n$ et $\frac{1}{2}$. Soit $\alpha \in \mathbb{R}^{+\star}$. On pose $Y_n = \frac{1}{2^n}\alpha^X$. Que vaut $\mathbb{E}(Y_n)$ ? Quelles sont les limites possibles de $\mathbb{E}(Y_n)$ lorsque $n \to +\infty$ ?

CORRECTION: On a

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{k=0}^{n} \frac{\alpha^k}{2^n} P(X=k) \\ &= \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} (\frac{1}{2})^k (\frac{1}{2})^{n-k} \alpha^k \\ &= \frac{1}{4^n} \sum_{k=0}^{n} \binom{n}{k} \alpha^k \\ &= \frac{(1+\alpha)^n}{4^n}\end{aligned}$$

en raison de la formule du binôme. Si $0 < \alpha < 3$ alors $0 < \frac{1+\alpha}{4} < 1$ et donc $\mathbb{E}(Y_n) \to 0$ lorsque $n \to +\infty$. Si $\alpha = 3$ alors $\frac{1+\alpha}{4} = 1$ et donc $\mathbb{E}(Y_n) = 1$ pour tout $n$. En particulier, $\mathbb{E}(Y_n) \to 1$ lorsque $n \to +\infty$. Si $\alpha > 3$ alors $\frac{1+\alpha}{4} > 1$ et alors $\mathbb{E}(Y_n) \to +\infty$ lorsque $n \to +\infty$.

**EXERCICE 36$^\star$:** Soit $X$ une variable aléatoire suivant une loi binomiale de paramètres $n$ et $p$. Pour quel $k$ la probabilité élémentaire $P(X=k)$ est-elle maximale ?

CORRECTION: On va étudier la croissance et décroissance des $P(X=k)$. On a

$$\frac{P(X=k+1)}{P(X=k)} = \frac{n-k}{k+1} \frac{p}{1-p} \ .$$

On a

$$\begin{aligned}&\frac{n-k}{k+1} \frac{p}{1-p} \geq 1 \\ &\Leftrightarrow (n-k)p \geq (1-p)(k+1) \\ &\Leftrightarrow np \geq k+1-p \\ &\Leftrightarrow k \leq (n+1)p - 1\end{aligned}$$

Soit $k_0 = E((n+1)p)$ la partie entière de $(n+1)p$. On en déduit que la "fonction" $k \to P(X=k)$ est croissante sur $[\![0, k_0]\!]$ et décroissante sur $[\![k_0, n]\!]$. Le maximum des $P(X=k)$ est atteint pour $k = k_0$.

EMMANUEL HEBEY, UNIVERSITÉ DE CERGY-PONTOISE, DÉPARTEMENT DE MATHÉMATIQUES, SITE DE SAINT-MARTIN, 2 AVENUE ADOLPHE CHAUVIN, 95302 CERGY-PONTOISE CEDEX, FRANCE