

Chapitre 16

Sommes de variables aléatoires et inégalités de concentration

Etant donnée une famille de variables aléatoires X_1, X_2, \dots, X_n avec $n \geq 1$, l'objectif de ce chapitre est de comprendre le comportement des quantités suivantes :

$$S_n = X_1 + X_2 + \dots + X_n \quad \text{et} \quad \bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n).$$

Remarque. \bar{X}_n est désignée (et notée maladroitement M_n dans vos livres) sous l'appellation « moyenne empirique ». Le symbole \sum est fréquemment employé pour alléger les notations impliquant une somme :

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{k=1}^n X_k.$$

La notation $\sum_{k=1}^n X_k$ indique ainsi que nous additionnons les termes X_k pour k allant de 1 jusqu'à n . L'indice k de sommation est muet, cela signifie (comme la variable d'intégration) que les quantités suivantes sont égales

$$\sum_{k=1}^n X_k = \sum_{j=1}^n X_j.$$

Exemple 16.0.1. 1. Supposons que nous mettions en place un schéma de Bernoulli avec une pièce équilibrée. A chaque étape, nous pourrions décider d'associer une variable aléatoire X_k (pour k allant de 1 à n) laquelle prend la valeur 1 lorsque le résultat est pile et la valeur -1 lorsque le résultat est face. Ainsi, le résultat (moyen) final de notre schéma de Bernoulli revient à considérer

$$\frac{1}{n}S_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} \sum_{k=1}^n X_k.$$

Notre intuition nous indique que nous devrions obtenir 0 dès que n est suffisamment grand car il devrait y avoir autant de piles que de faces puisque la pièce est équilibrée. Ce chapitre va nous permettre de formaliser et quantifier tout ceci.

2. Voyons un deuxième exemple issu d'un domaine n'ayant a priori peu de rapport avec le chapitre. Lors d'une élection, imaginons que nous disposions de n électeurs¹ X_1, \dots, X_n et de 2 candidats (l'un que nous désignons par $+1$ et l'autre par -1). Chaque électeur indique s'il vote pour le candidat $+1$ ou pour le candidat -1 . Si le scrutin s'effectue de sorte que le candidat élu soit celui qui a obtenu la majorité de votes, il suffit de regarder le signe de la quantité suivante :

$$S_n = X_1 + \dots + X_n.$$

En effet, si $S_n > 0$, cela signifie qu'il y a eu plus de $+1$ que de -1 et le candidat $+1$ est élu ; si $S_n < 0$, le gagnant sera le candidat -1 .

Remarque. Dans une dictature, le résultat du scrutin correspondra, par exemple, au vote X_1 de l'électeur numéro 1 (en supposant qu'il s'agisse du dictateur lui-même). Notons, qu'il est aussi possible de modéliser des scrutins de majorités récursives comme c'est le cas aux élections présidentielles aux Etats-Unis.

Nous focalisons le début de notre étude sur S_n et reviendrons ultérieurement sur l'étude du comportement de \bar{X}_n .

1. En pratique, il est supposé que n est suffisamment grand, de sorte qu'il ne puisse y avoir d'égalité après le décompte des votes.

16.1 Loi d'une famille de variables aléatoires

Puisque les quantités S_n et $\overline{X_n}$ impliquent une famille de variables aléatoires, il est naturel de savoir de quelle manière ces variables aléatoires déterminent la loi de S_n . Commençons donc par l'étude d'un couple de variables aléatoires.

16.1.1 Loi d'un couple $(X; Y)$

Débutons par un exemple.

Exemple 16.1.1. En période de réglage de ses machines, le contrôle qualité d'une usine effectue de nombreux prélèvements dans la production et relève, pour chaque pièce, si elle a un défaut A ou un défaut B . Nous notons X (resp. Y) la variable aléatoire qui compte le nombre de défaut A (resp. B). Nous supposons que la variable $Y \in \{0; 1; 2\}$ et la variable $X \in \{0; 1\}$ et nous avons à disposition la loi jointe du couple (X, Y) , il s'agit d'un tableau à double entrée² :

	Y	0	1	2
X		0.07	0.18	0.17
	0			
	1	0.15	0.20	0.23

Ce tableau indique, par exemple, que $\mathbb{P}((X; Y) = (0; 1)) = 0.18$. Si maintenant $S = X + Y$ est la variable aléatoire qui compte le nombre défaut (de type A et de type B) d'une pièce, nous savons que $S \in \{0; 1; 2; 3\}$ mais il reste à savoir avec quelles probabilités ces valeurs sont prises. Pour cela, nous devons utiliser la loi du couple $(X; Y)$. Pour la valeur 0, c'est assez simple :

$$\mathbb{P}(S = 0) = \mathbb{P}((X; Y) = (0; 0)) = 0.07.$$

En revanche, pour la valeur 1, nous avons deux possibilités : soit nous avons un défaut de type A soit un défaut de type B . Ainsi,

$$\mathbb{P}(S = 1) = \mathbb{P}((X; Y) = (1; 0)) + \mathbb{P}((X; Y) = (0; 1)) = 0.15 + 0.18 = 0.33$$

Et nous pourrions procéder de même pour les valeurs 2 et 3. En résumé, la loi de S est donnée dans le tableau suivant :

s	0	1	2	3
$\mathbb{P}(S = s)$	0.07	0.33	0.37	0.23

Remarque. 1. En fait, le procédé décrit ci-dessus permet de déterminer la loi de X et celle de Y (les lois marginales) à partir de la loi du couple $(X; Y)$. En effet : si $X = 0$ alors Y peut prendre les valeurs 0, 1 ou 2 d'où

$$\mathbb{P}(X = 0) = \mathbb{P}((X; Y) = (0; 0)) + \mathbb{P}((X; Y) = (0; 1)) + \mathbb{P}((X; Y) = (0; 2)) = 0.07 + 0.18 + 0.17 = 0.42$$

ce qui revient à additionner les probabilités de la première ligne du tableau exposant la loi du couple. De la même manière,

$$\mathbb{P}(X = 1) = \mathbb{P}((X; Y) = (1; 0)) + \mathbb{P}((X; Y) = (1; 1)) + \mathbb{P}((X; Y) = (1; 2)) = 0.15 + 0.20 + 0.23 = 0.58$$

grâce à la deuxième ligne. En résumé,

i	0	1
$\mathbb{P}(X = i)$	0.42	0.58

Par les mêmes arguments, nous trouvons la loi de Y à partir de la loi du couple $(X; Y)$ (cette fois-ci en additionnant les probabilités, colonnes par colonnes) :

j	0	1	2
$\mathbb{P}(Y = j)$	0.22	0.38	0.40

2. Lequel indique ce que vaut $\mathbb{P}((X; Y) = (i; j))$ lorsque $i \in \{0; 1\}$ et $j \in \{0; 1; 2\}$.

2. La réciproque de l'observation précédente est fautive : en général, les lois des marginales X et Y ne déterminent pas la loi du couple $(X; Y)$. Voyons cela sur un exemple. Supposons que les lois de X et de Y soient données par

i	-1	0	1
$\mathbb{P}(X = i)$	0.24	0.5	0.25

j	-1	0	1
$\mathbb{P}(Y = j)$	0.24	0.5	0.25

Nous constatons que les deux lois jointes données ci-dessous fournissent les mêmes lois marginales :

$X \backslash Y$	-1	0	1
-1	0	0.25	0
0	0.25	0	0.25
1	0	0.25	0

$X \backslash Y$	0	1	2
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

Exercice à traiter : 4 page 107 et 44 page 418 (cf. méthode 2 page 407).

Nous venons d'observer que la connaissance des lois marginales ne suffit pas pour déterminer la loi d'un couple de variables aléatoires. L'ajout d'une hypothèse d'indépendance entre X et Y permet de régler ce problème et simplifie grandement les calculs.

16.1.2 Loi de variables indépendantes

Lorsque les variables X_1, X_2, \dots, X_n sont deux à deux indépendantes, il est beaucoup plus aisé de déterminer la loi du vecteur (X_1, X_2, \dots, X_n) puisqu'il suffit de multiplier les lois marginales.

Exemple 16.1.2. Considérons deux dés équilibrés à 4 faces. Notons X_1 le résultat du premier dé et X_2 le résultat du deuxième, les résultats obtenus sont indépendants et nous sommes dans une situation d'équiprobabilité : pour $i \in \{1; 2\}$, nous avons

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 2) = \mathbb{P}(X_i = 3) = \mathbb{P}(X_i = 4) = \frac{1}{4}.$$

Nous en déduisons alors la loi du couple (X_1, X_2) :

$X_1 \backslash X_2$	1	2	3	4
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Laquelle nous permet ensuite, par exemple, de déterminer la loi de la somme $S = X_1 + X_2$:

s	2	3	4	5	6	7	8
$\mathbb{P}(S = s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Notons que l'hypothèse d'indépendance facilite grandement nos calculs. Par exemple,

$$\begin{aligned} \mathbb{P}(S = 4) &= \mathbb{P}((X_1; X_2) = (1; 3)) + \mathbb{P}((X_1; X_2) = (3; 1)) + \mathbb{P}((X_1; X_2) = (2; 2)) \\ &= \mathbb{P}(X_1 = 1) \times \mathbb{P}(X_2 = 3) + \mathbb{P}(X_1 = 3) \times \mathbb{P}(X_2 = 1) + \mathbb{P}(X_1 = 2) \times \mathbb{P}(X_2 = 2) \\ &= \frac{3}{16}. \end{aligned}$$

Exercice à traiter : 45 et 46 page 418.

Voyons à présent une définition.

Définition 16.1.1. Un échantillon de taille n est un vecteur aléatoire $(X_1; X_2; \dots; X_n)$ dont les coordonnées sont mutuellement indépendantes et de même loi.

Remarque. Comme observé dans l'exemple ci-dessus, la loi du vecteur $(X_1; X_2; \dots; X_n)$ s'obtient en multipliant les lois marginales. De plus, les variables étant de même loi, nous avons

$$\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n] \quad \text{et} \quad \text{Var}(X_1) = \dots = \text{Var}(X_n).$$

Ceci sera utilisé à de nombreuses reprises.

Puisque nous avons étudié la loi de Bernoulli et la binomiale en début d'année, il semble naturel de se poser les questions suivantes :

1. Si $(X_1; X_2; \dots; X_n)$ est un échantillon de taille n obtenu à partir d'une loi de Bernoulli de paramètre $p \in [0; 1]$ (i.e. $X_k \sim B_e(p)$ pour tout $k = 1, \dots, n$). Que pouvons-nous dire de la loi de

$$S_n = X_1 + X_2 + \dots + X_n \quad ?$$

2. Si $(X_1; X_2; \dots; X_m)$ est un échantillon de taille m obtenu à partir d'une loi Binomiale de paramètre $p \in [0; 1]$ et $n \geq 1$ (i.e. $X_k \sim B(n; p)$ pour tout $k = 1, \dots, m$). Que pouvons-nous dire de la loi de

$$S_m = X_1 + X_2 + \dots + X_m \quad ?$$

Bien que la deuxième question soit à notre portée, nous allons nous focaliser uniquement sur la première.

Proposition 96. Si $(X_1; X_2; \dots; X_n)$ est un échantillon de taille n obtenu à partir d'une loi de Bernoulli de paramètre $p \in [0; 1]$ (i.e. $X_k \sim B_e(p)$ pour tout $k = 1, \dots, n$) alors

$$S_n = X_1 + X_2 + \dots + X_n \quad \text{suit une loi binomiale de paramètre } n \text{ et } p.$$

Autrement dit $S_n \sim B(n; p)$.

Exercices à traiter : 5, 6 page 409 et 50-52-54 page 419 (cf. Méthode 3 page 409).

16.2 Espérance de variables aléatoires

Comme cela a été étudié en classe de première ou de seconde, il est souvent nécessaire de chercher à résumer des informations probabilistes (ou statistiques) à l'aide de paramètres. L'espérance d'une variable aléatoire a été rencontré en classe de première (nous rappelons la définition ci-dessous), il est alors naturel de chercher à comprendre comment l'espérance se comporte vis-à-vis des sommes de variables aléatoires. Autrement dit, si $S_n = X_1 + \dots + X_n$ que vaut

$$\mathbb{E}[S_n] = \mathbb{E}[X_1 + \dots + X_n] \quad ?$$

Définition 16.2.1. Si X est une variable aléatoire discrète, que x_1, \dots, x_k sont les k valeurs prises par X avec probabilités $p_i = \mathbb{P}(X = x_i)$ pour tout $i \in \{1, \dots, k\}$ alors

$$\mathbb{E}[X] = \sum_{i=1}^k x_i p_i.$$

Cette définition permet d'obtenir la proposition suivante, laquelle indique que l'espérance vérifie des propriétés de linéarité (comme l'intégrale finalement)

Proposition 97. Soient X et Y deux variables aléatoires et $a \in \mathbb{R}$. Alors

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- $\mathbb{E}[aX] = a\mathbb{E}[X]$.

Démonstration. La démonstration est laissée en exercice et constitue un entraînement intéressant de manipulation de sommes. □

En application de ceci, nous pouvons retrouver un résultat énoncé plus tôt dans l'année.

Exemple 16.2.1. Soit $(X_1; X_2; \dots; X_n)$ est un échantillon de taille n obtenu à partir d'une loi de Bernoulli de paramètre $p \in [0; 1]$ (i.e. $X_k \sim B_e(p)$ pour tout $k = 1, \dots, n$). En particulier,

$$\mathbb{E}[X_k] = p \quad \text{pour tout } k = 1, \dots, n.$$

En conséquence, d'après la proposition précédente

$$\mathbb{E}[S_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] = n \times p.$$

Puisque $S_n \sim B(n; p)$, nous retrouvons bien le fait que l'espérance d'une loi binomiale (la loi suivie par S_n) de paramètres n et p vaut $n \times p$.

Exercices à traiter : 35 et 36 page 418 (cf. méthode 1 page 407).

L'utilisation de la linéarité de l'espérance nous permet de constater la chose suivante : si (X_1, \dots, X_n) est un échantillon de taille n (i.e. les variables sont mutuellement indépendantes et de même loi) alors

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n}(\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]) = \mathbb{E}[X_1]$$

puisque $\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n]$ comme les variables aléatoires sont de même loi. Il est donc naturel de se demander si $\mathbb{E}[X_1]$ est une bonne manière de résumer le comportement de \bar{X}_n . C'est l'objet de la section suivante.

16.2.1 Moyenne empirique et loi forte des grands nombres

Dans cette section nous allons expliquer pour quelles raisons $\mathbb{E}[X_1]$ fournit des informations pertinentes sur la moyenne empirique \bar{X}_n et cela nous permettra de justifier notre intuition quant à l'exemple impliquant une série de piles ou faces avec une pièce équilibrée (au bout d'un moment, il y a autant de piles que de faces).

Théorème 98 (Loi forte des grands nombres). Soient (X_1, X_2, \dots, X_n) un échantillon de taille n . Nous avons alors l'équivalence suivante :

$$\mathbb{E}[X_1] < +\infty \quad \Longleftrightarrow \quad \lim_{n \rightarrow +\infty} \bar{X}_n = \mathbb{E}[X_1].$$

Remarque. Pour être précis, la convergence de \bar{X}_n a lieu « presque sûrement ». Ceci signifie qu'il existe un sous-ensemble Ω' de l'univers Ω (sous-jacent à l'expérience aléatoire liée à l'échantillon) tel que $\mathbb{P}(\Omega') = 1$ et

$$\lim_{n \rightarrow +\infty} \bar{X}_n(\omega) = \mathbb{E}[X_1] \quad \text{pour tout } \omega \in \Omega'.$$

Ce théorème nous permet de reprendre notre exemple initial de piles ou faces.

Exemple 16.2.2. Considérons un échantillon de taille n obtenu à partir d'une variable X_1 dont la loi est donnée par

i	-1	1
$\mathbb{P}(X_1 = i)$	$\frac{1}{2}$	$\frac{1}{2}$

Cet échantillon correspond aux résultats de nos piles (-1) ou faces ($+1$) obtenus à partir d'une pièce équilibrée. Dans ce cas, la loi forte des grands nombres nous assure que

$$\lim_{n \rightarrow +\infty} \bar{X}_n = \lim_{n \rightarrow +\infty} \frac{1}{n}(X_1 + \dots + X_n) = \mathbb{E}[X_1].$$

Or, un rapide calcul nous montre que $\mathbb{E}[X_1] = 0$. Nous avons donc, lorsque n est suffisamment grand, autant de piles que de faces.

Exercices à traiter : 14,13 page 413 (cf. méthode 7 page 413).

La loi forte des grands nombres est un résultat théorique intéressant mais peu utile en pratique car il ne précise pas à partir de quand (i.e. à quel point n doit être grand) pour que la moyenne empirique soit proche de la moyenne $\mathbb{E}[X_1]$. Pour quantifier cela, nous allons devoir introduire la notion d'inégalités de concentration et utiliser la variance de l'échantillon.

16.3 Précisions sur la loi forte des grands nombres

Débutons par étudier des notions liées à la variance de variables aléatoires avant d'aborder les inégalités de concentration.

16.3.1 Variance de variables indépendantes

Définition 16.3.1. Si X est une variable aléatoire discrète, que x_1, \dots, x_k sont les k valeurs prises par X avec probabilités $p_i = \mathbb{P}(X = x_i)$ pour tout $i \in \{1, \dots, k\}$ alors

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad \text{avec} \quad \mathbb{E}[X^2] = \sum_{i=1}^k x_i^2 p_i.$$

L'écart-type associé vaut alors $\sigma(X) = \sqrt{\text{Var}(X)}$.

Remarque. 1. En fait, la formule ci-dessus est déjà un résultat³. Si $m = \mathbb{E}[X]$ alors la véritable définition de la variance est :

$$\text{Var}(X) = \mathbb{E}[(X - m)^2] = \sum_{i=1}^k (x_i - m)^2 p_i.$$

Il suffit ensuite de développer le carré et la somme pour montrer que $\mathbb{E}[(X - m)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

2. Il est aussi possible de voir la variance comme le minimum de la fonction $\phi(x) = \mathbb{E}[(X - x)^2]$ avec $m \in \mathbb{R}$. En effet, $\phi(m)$ s'écrit aussi

$$\phi(x) = \mathbb{E}[X^2] - 2x\mathbb{E}[X] + x^2.$$

Une étude de variation montre alors que ϕ atteint son minimum en $x = \mathbb{E}[X]$:

$$\text{Var}(X) = \phi(\mathbb{E}[X]) \leq \phi(x) = \mathbb{E}[(X - x)^2] \quad \text{pour tout } x \in \mathbb{R}.$$

Contrairement à l'espérance, la variance n'est pas linéaire (à cause de la présence du carré).

Proposition 99. Soit X une variable aléatoire et $a, b \in \mathbb{R}$ alors

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

En général, sans hypothèses supplémentaires, la relation suivante n'est pas vérifiée

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y).$$

En revanche, si X et Y sont indépendantes, les choses sont plus simples.

Proposition 100. Soient X et Y deux variables aléatoires indépendantes alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y). \quad (16.3.1)$$

Remarque. 1. Il se trouve que ce résultat peut s'obtenir sous une hypothèse plus faible. Pour cela, introduisons la notion de corrélation entre deux variables aléatoires. Ceci se mesure via la covariance qui est définie comme suit

$$\text{Cov}(X; Y) = \mathbb{E} \left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) \right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Nous dirons que X et Y sont non corrélées si $\text{Cov}(X; Y) = 0$; en particulier, des variables indépendantes sont non-corrélées puisque l'indépendance implique que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, la réciproque est fautive cependant. De manière générale, nous avons (après un calcul) l'identité suivante :

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X; Y) + \text{Var}(Y). \quad (16.3.2)$$

Ainsi, l'identité (16.3.1) reste valide lorsque X et Y sont non-corrélées.

3. Il s'agit de la formule de Koenig-Huygens.

2. Pour simplifier, supposons que les variables X et Y soient centrées (i.e. $\mathbb{E}[X] = \mathbb{E}[Y] = 0$). Dans ce cas, nous avons

$$\text{Cov}(X; Y) = \mathbb{E}[XY].$$

Il est alors facile de vérifier que la covariance est une forme bilinéaire, symétrique et définie positive. En d'autres termes, la covariance vérifie les mêmes propriétés algébriques que le produit scalaire que nous avons étudié dans le plan et dans l'espace en classe de première et de terminale. Ceci signifie donc que la notion de non corrélation s'apparente à une forme d'orthogonalité entre des variables aléatoires. Puisque nous pouvons définir une notion d'orthogonalité entre des variables aléatoires, cela signifie que tout ce que nous avons étudié cette année dans l'espace est valable dans ce nouveau contexte. Ainsi, grâce à ce point de vue géométrique, il est possible de montrer que $\mathbb{E}[X]$ correspond à une projection orthogonale de la variable X sur un sous-espace particulier. Nous n'entrerons pas plus dans le détail de ce genre de considérations qui sont, bien entendu, hors programme.

Voyons une application de la proposition précédente.

Exemple 16.3.1. Si $X \sim B(n; p)$ alors $X = X_1 + \dots + X_n$ avec X_1, \dots, X_n des variables aléatoires et indépendantes de même loi telles que $X_1 \sim B_e(p)$. Dans ce cas, grâce à l'indépendance

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = np(1-p)$$

puisque $\text{Var}(X_1) = p(1-p)$. Ceci montre aussi que $\sigma(X) = \sqrt{np(1-p)}$.

Exercice à traiter : 39 page 418 et 52 page 419.

Voyons un deuxième exemple de ceci.

Exemple 16.3.2. Un industriel fabrique des plats en verre qui doivent résister à de fortes températures afin de pouvoir être utilisés dans un four de cuisson. Pour vérifier la résistance du plat, on le soumet à une température de 350°C . Il a été constaté qu'en moyenne, sur un grand nombre de plats testés sortant de l'usine, 1,5% des plats ne supportent pas une telle température et cassent. Choisissons au hasard 200 plats produits par l'usine et effectuons, pour chacun d'eux, le test de résistance. Etant donné le grand nombre de plats produits par l'usine, ce choix peut être assimilé à un tirage fait de façon indépendante et avec remise.

Désignons par R la variable aléatoire comptant le nombre de plats résistants au test.

1. Calculons $\mathbb{E}[R]$ et $\sigma(R)$.

Soit X la variable aléatoire égale à 1 si le plat résiste au test et 0 dans le cas contraire. X suit une loi de Bernoulli de paramètre $p = 0.985$. En effet, d'après l'énoncé, nous savons que $\mathbb{E}[X] = 1 - 0.015 = 0,985$. Or, puisque lorsque $X \sim B_e(p)$ nous avons $\mathbb{E}[X] = p$ d'où le résultat. Par conséquent, $\text{Var}(X) = p(1-p) = 0.014775$. Considérons alors un échantillon (X_1, \dots, X_n) de taille $n = 200$ chacune des variables aléatoires suivant une loi de Bernoulli $B_e(0.985)$. En outre,

$$R = X_1 + \dots + X_n,$$

c'est pourquoi

$$\mathbb{E}[R] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = 200 \times 0.985 = 197.$$

De plus, puisque les variables X_1, X_2, \dots, X_n sont mutuellement indépendantes, nous avons aussi

$$\text{Var}(R) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = 200 \times 0.014775 \approx 2,995$$

d'où $\sigma(R) \approx \sqrt{2.995} \approx 1.72$.

2. Interprétons les résultats obtenus.

En moyenne, sur un grand nombre d'échantillons de 200 plats, nous pouvons espérer trouver 197 plats résistants avec un écart-type proche de 1,72.

Remarque. Nous aurions pu utiliser le fait que $X \sim B(200; 0.985)$ pour traiter la première partie de l'exercice.

16.3.2 Inégalités de concentration

La loi forte des grands nombres (cf. théorème 16.2.2) nous assure que sur un ensemble de probabilité 1, nous avons

$$\lim_{n \rightarrow +\infty} \bar{X}_n = \lim_{n \rightarrow +\infty} \frac{1}{n}(X_1 + \dots + X_n) = \mathbb{E}[X_1]$$

lorsque (X_1, \dots, X_n) désigne un échantillon de taille n . Nous en déduisons que la probabilité que \bar{X}_n ne converge pas vers $\mathbb{E}[X_1]$ doit tendre vers 0 lorsque $n \rightarrow +\infty$. Étant donné $t > 0$ et $n \geq 1$, il est donc pertinent de quantifier (en fonction de t) à quelle vitesse la probabilité⁴

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| \geq t)$$

tend vers 0. C'est le contenu de l'inégalité de concentration suivante.

Proposition 101 (Inégalité de Bienaymé-Tchebychev). *Soit (X_1, \dots, X_n) un échantillon⁵ de taille n alors, pour tout $t > 0$, nous avons*

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| \geq t) \leq \frac{\text{Var}(X_1)}{nt^2}. \quad (16.3.3)$$

Démonstration. La démonstration repose sur une variante de l'inégalité de Markov qui nous assure que, pour tout $t > 0$ et toute variable aléatoire X (de variance finie), nous avons

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}. \quad (16.3.4)$$

Il suffit ensuite d'appliquer cette inégalité à $X = \bar{X}_n$ en observant que $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1]$ (par linéarité de l'espérance) et $\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_1)}{n}$ puisque les variables X_1, \dots, X_n sont indépendantes et de même loi. \square

Remarque. 1. Bien entendu, la majoration n'a d'intérêt que si le membre de droite de l'inégalité de Bienaymé-Tchebychev est inférieure ou égale à 1.

2. L'entier n étant fixé, l'inégalité (16.3.3) indique que la probabilité que \bar{X}_n s'écarte de la moyenne $\mathbb{E}[X_1]$ d'au moins t devient de plus en plus petit lorsque la valeur de t augmente.
3. L'assertion précédente permet alors d'affirmer que la probabilité du complémentaire devient de plus en plus proche de 1 lorsque t augmente. Autrement, dès que t est suffisamment grand, l'évènement

$$\bar{X}_n \in]\mathbb{E}[X_1] - t; \mathbb{E}[X_1] + t[$$

est quasiment certain : la moyenne empirique s'est concentrée dans l'intervalle $[\mathbb{E}[X_1] - t; \mathbb{E}[X_1] + t]$. Plus précisément, ceci est visible via (16.3.4) (il suffit ensuite de remplacer X par \bar{X}_n). Si $\delta > 0$ est un seuil donné, il est possible de trouver t de sorte que

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| \geq t) \leq \delta.$$

Pour cela, il suffit de constater que $\frac{\text{Var}(X_1)}{nt^2} = \delta \iff t = \sqrt{\frac{\text{Var}(X_1)}{\delta}}$. Ainsi, avec une probabilité supérieure à $1 - \delta$ nous avons⁶

$$|\bar{X}_n - \mathbb{E}[X_1]| < t = \sqrt{\frac{\text{Var}(X_1)}{\delta}} \iff \bar{X}_n \in \left] \mathbb{E}[X_1] - \sqrt{\frac{\text{Var}(X_1)}{\delta}}; \mathbb{E}[X_1] + \sqrt{\frac{\text{Var}(X_1)}{\delta}} \right[$$

et nous constatons que la proximité de \bar{X}_n par rapport à $\mathbb{E}[X_1]$ dépend de sa variance et du seuil δ . Ce genre de raisonnement sera repris en exercice.

4. Il faut noter que la vitesse de décroissance (en fonction de t) du membre de droite (dans l'inégalité de Bienaymé-Tchebychev) est assez lente et est souvent sous-optimale. Dans le cas d'une somme de variable aléatoire de Bernoulli, nous devrions avoir une décroissance impliquant une fonction de la forme $t \mapsto e^{-\frac{t^2}{2}}$ laquelle tend beaucoup plus vite vers 0 que la fonction $t \mapsto \frac{1}{t^2}$ donnée par l'inégalité de Bienaymé-Tchebychev.

4. Laquelle correspond à la probabilité que \bar{X}_n s'écarte d'au moins t de $\mathbb{E}[X_1]$.

5. Rappelons que X_1, \dots, X_n sont donc des variables aléatoires mutuellement indépendantes et de même loi.

6. Le raisonnement suivant sera utilisé à de nombreuses reprises : soit A un évènement tel que $\mathbb{P}(A) \leq \delta$ alors $\mathbb{P}(A^c) > 1 - \delta$.

Voyons deux applications de tout cela.

Exemple 16.3.3. Dans une société de démarchage par téléphone, nous estimons que 40% des personnes appelées répondent effectivement. Si n personnes sont appelées, notons X_k la variable aléatoire qui vaut 1 si la k -ième personne répond et vaut 0 sinon. Il est supposé que les variables X_1, \dots, X_n sont mutuellement indépendantes de même loi (elles suivent une loi de Bernoulli de paramètre 0.4). En particulier, elles ont la même espérance et la même variance :

$$\mathbb{E}[X_1] = 0.4 \quad \text{et} \quad \text{Var}(X_1) = 0.24.$$

Alors, d'après l'inégalité de Bienaymé-Tchebychev (16.3.3), nous avons

$$\mathbb{P}(|\bar{X}_n - 0.4| \geq t) \leq \frac{0.24}{nt^2} \quad \text{pour tout } t > 0.$$

En particulier, si $t = 0.1$ (cette valeur correspond à une précision souhaitée) nous avons

$$\mathbb{P}(|\bar{X}_n - 0.4| \geq 0.1) \leq \frac{24}{n}.$$

Ceci permet alors de déterminer le nombre n de personnes nécessaire pour que la probabilité $\mathbb{P}(|\bar{X}_n - 0.4| \geq 0.1)$ soit plus petite qu'un risque donné. Par exemple, si le risque vaut 0.024 nous observons qu'il faut choisir $n = 1\,000$. Ceci s'interprète comme suit : en appelant $n = 1\,000$ personnes, la probabilité que le nombre de personnes qui répondent se trouve en dehors de l'intervalle $]300; 500[$ est inférieur à 0.024. Autrement dit, il est quasiment certains qu'entre 300 et 500 personnes répondent à leur téléphone.

Remarque. Cet exemple illustre bien l'importance des inégalités de concentration afin d'obtenir des intervalles de fluctuations (ou des intervalles de confiance).

Le deuxième exemple d'application est plus théorique et permet de démontrer la loi faible des grands nombres

Théorème 102 (Loi faible des faibles des grands nombres). *Soit (X_1, \dots, X_n) un échantillon de taille n tel que la variance commune soit finie $\text{Var}(X_1)$. Alors, pour tout $t > 0$, nous avons*

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(|\bar{X}_n - \mathbb{E}[X_1]| \geq t\right) = 0$$

où, rappelons le, $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$.

Remarque. Ce résultat indique que la probabilité que la moyenne empirique s'écarte d'au moins t de $\mathbb{E}[X_1]$ tend vers 0 lorsque $n \rightarrow +\infty$.

Démonstration. La démonstration est immédiate grâce à l'inégalité de Bienaymé-Tchebychev (16.3.3) dans laquelle il suffit de faire tendre $n \rightarrow +\infty$. \square

Exercices à traiter : 9 page 411 (cf. méthode 5 page 411); 11 page 413 (cf. méthode 6 page 413); 17/18 page 415 puis 114-116 page 427 (cf. méthode 9 page 415); 118 page 427.

16.4 Pour aller plus loin

Voici quelques éléments de réflexion permettant d'aller plus loin que cours qui vient d'être exposé. Bien que simplifiés, certains passages peuvent sembler plus ardues à lire sans prendre une feuille et un crayon. Les sections sont plus ou moins indépendantes les unes des autres et le lecteur pourra les parcourir selon ses envies.

16.4.1 Analyse Booléenne et démocratie

Au début de ce chapitre nous avons vu que les probabilités permettaient de modéliser certains mode de scrutin. Rappelons qu'il s'agissait de représenter le scrutin utilisé pour agglomérer les n votes des électeurs pour le candidat +1 ou le candidat -1.

Un mode de scrutin étant donné (i.e. une fonction f dans laquelle les votes des électeurs sont placés et dont le résultat indique lequel des candidats +1 ou -1 est le vainqueur), il est naturel de s'interroger sur l'influence

7. Lequel s'obtient en observant que $\left\{|\bar{X}_n - 0.4| \geq 0.1\right\}^c = \left\{|\bar{X}_n - 0.4| < 0.1\right\} = \left\{\left|\frac{1}{1000} \sum_{k=1}^n X_k - 0.4\right| < 0.1\right\}$ ce qui signifie que $\sum_{k=1}^n X_k \in](0.4 - 0.1) \times 1000; (0.4 + 0.1) \times 1000[=]300; 500[$.

du i -ème votant. En particulier, est-il possible que la modification du i -ème vote change le résultat de l'élection ?

Etant donnée X_1, \dots, X_n correspondant aux votes des électeurs, nous pouvons définir l'influence de la i ème coordonnée par

$$\mathbb{P}(f(X_1, \dots, X_n) \neq f(X_1, \dots, \tilde{X}_i, \dots, X_n))$$

où \tilde{X}_i est le vote modifiée du i ème électeur.

Il existe toute une série de résultats liés à ce genre de notions :

- quel est le mode de scrutin (i.e. la fonction f) le plus stable ? Autrement dit, quel est le scrutin le moins sensible aux changements de vote ?
- étant donné un mode de scrutin, est-il possible d'estimer l'influence des coordonnées ?
- pourquoi ne pas étudier l'influence d'un couple de votants ?
- ...

Les inégalité de concentration et les résultats qui y sont liés sont des outils très pratiques pour répondre à ce genre de problématiques

16.4.2 Inégalités de concentration gaussienne

Il se trouve que la vitesse de décroissance donnée par l'inégalité de Bienaymé-Tchebychev est sous-optimale dans de nombreux cas (notamment pour la loi binomiale). Voyons un autre type de loi pour laquelle, il est possible de faire mieux. Une variable aléatoire normale, centrée et réduite X ($X \sim \mathcal{N}(0; 1)$ en abrégé) est une variable aléatoire telle que

$$\mathbb{P}(a \leq X \leq b) = \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

Nous allons montrer⁸ que, pour tout $t > 0$,

$$\mathbb{P}(X \geq t) \leq e^{-\frac{t^2}{2}}.$$

Le théorème de croissances comparées nous assure que la vitesse de décroissance de $t \mapsto e^{-\frac{t^2}{2}}$ est bien supérieure à celle donnée par l'inégalité de Bienaymé-Tchebychev (en $t \mapsto \frac{1}{t^2}$). Pour obtenir ce résultat, nous allons utiliser deux nouvelles notions :

1. celle de fonction indicatrice d'un ensemble $A \subset \Omega$ (avec Ω l'univers). Etant donné A , nous définissons la fonction $\omega \mapsto 1_A(\omega)$ par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

De plus, $\mathbb{E}[1_A] = \mathbb{P}(A)$.

2. L'espérance est monotone : si $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Observons alors le fait suivant, étant donné $\lambda > 0$, nous avons

$$1_{X(\omega) \geq t} \leq e^{-\lambda t} e^{\lambda X(\omega)} \quad \text{pour tout } t \geq 0 \quad \text{et tout } \omega \in \Omega. \quad (16.4.1)$$

En effet, posons $A = \{\omega \in \Omega ; X(\omega) \leq t\}$ et observons les deux de figures suivants :

- $\omega \notin A$ donc $1_A(\omega) = 0$ et l'inégalité est évidente puisque $e^x > 0$ pour tout $x \in \mathbb{R}$.
- $\omega \in A$ donc $1_A(\omega) = 1$ et $X(\omega) - t \geq 0$. En conséquence,

$$\lambda X(\omega) - \lambda t \geq 0 \quad \text{d'où} \quad e^{\lambda X(\omega) - \lambda t} \geq e^0 = 1 = 1_A(\omega)$$

puisque $x \mapsto e^x$ est une fonction croissante.

8. Pour simplifier (car certaines propriétés utilisées ne sont pas connues du lecteur), nous allons seulement établir une estimation de la probabilité que X s'écarte d'au moins t de sa moyenne. Toutefois, les idées impliquées sont précisément celles qu'il est possible d'employer pour obtenir une inégalité de concentration dans un tel contexte. Le résultat que nous allons établir peut s'obtenir avec des arguments plus élémentaires (mais dont la mise en oeuvre n'est plus possible dans un contexte plus élaboré) que ceux exposés dans ce cours.

Appliquons alors l'espérance à l'inégalité (16.4.1). Ceci entraîne alors que

$$\mathbb{E}[1_{X(\omega) \geq t}] \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X(\omega)}] \iff \mathbb{P}(X \geq t) \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X}]$$

puisque $\mathbb{E}[1_A] = \mathbb{P}(A)$ pour tout évènement A . Pour conclure, il ne nous reste plus qu'à déterminer $\phi : \lambda \mapsto \mathbb{E}[e^{\lambda X}]$ et à choisir le meilleur λ possible.

Débutons par déterminer ϕ . Il se trouve que la fonction ϕ vérifie l'équation différentielle suivante :

$$y'(\lambda) = \lambda y(\lambda) \quad \text{avec} \quad y(0) = 1.$$

Cette équation différentielle est assez simple à résoudre : puisque $\phi(\lambda) > 0$ pour tout $\lambda > 0$, nous avons

$$\phi' = \lambda \phi \iff \frac{\phi'}{\phi} = \lambda.$$

En conséquence, en prenant une primitive (en la variable λ) dans chaque membre, nous en déduisons que

$$\ln \phi(\lambda) = \frac{\lambda^2}{2} + C \quad \text{avec} \quad C \in \mathbb{R}.$$

D'où (en passant à l'exponentielle),

$$\phi(\lambda) = e^{\frac{\lambda^2}{2}} \times e^C.$$

La condition initiale $y(0) = 1$ montre que $C = 0$, nous avons donc établi, lorsque $X \sim \mathcal{N}(0; 1)$, que $\mathbb{E}[e^{\lambda X}] = e^{\frac{\lambda^2}{2}}$.

Nous pouvons à présent conclure en optimisant en λ . Nous venons de montrer que

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t} e^{\frac{\lambda^2}{2}} \quad \text{pour tout} \quad t \geq 0 \quad \text{et} \quad \lambda > 0.$$

Il suffit ensuite d'observer (en étudiant les variations de la fonction $\psi(\lambda) = -\lambda t + \frac{\lambda^2}{2}$) que

$$\min_{\lambda \in \mathbb{R}_+^*} e^{-\lambda t} e^{\frac{\lambda^2}{2}} = e^{-\frac{t^2}{2}} \quad \text{pour tout} \quad t \geq 0.$$

Nous avons bien établi que $\mathbb{P}(X \geq t) \leq e^{-\frac{t^2}{2}}$ pour tout $t \geq 0$.

Remarque. Bien que la démonstration peut sembler un peu longue, il faut tout de même souligner que l'argument principal consiste à résoudre une équation différentielle. Ce procédé est assez systématique pour obtenir des inégalités de concentrations et les équations (ou inéquations) différentielles sous-jacentes (dépendant de la loi de la variable aléatoire étudiée) peuvent s'obtenir de différentes façons. L'une d'entre elles, très féconde, repose sur des inégalités dite « fonctionnelles » (celle de Poincaré ou de Sobolev logarithmique pour ne citer qu'elles) et permettent d'obtenir des résultats remarquables sans avoir à déployer des arguments techniques plus élaborés que ceux que nous venons d'exposer. Nous n'irons pas plus loin dans cette aparté qui nous a déjà fait explorer des sentiers se trouvant déjà très éloignés du programme.

16.4.3 Théorèmes limites

Etant donné un échantillon (X_1, X_2, \dots, X_n) dont l'espérance commune est finie, la loi forte des grands nombres nous indique que la somme $S_n = X_1 + \dots + X_n$ converge vers $\mathbb{E}[X_1]$ à condition de renormaliser S_n par un facteur n : autrement dit

$$\lim_{n \rightarrow +\infty} \frac{1}{n} S_n = \mathbb{E}[X_1].$$

Nous sommes en droit de nous demander ce qu'il peut se produire si nous avons renormalisé par $n^{1/p}$ avec $p \in]0; 2[$? Autrement dit, si $p \in]0; 2[$ que dire⁹ de

$$\lim_{n \rightarrow +\infty} \frac{1}{n^{1/p}} S_n = ?$$

Lorsque $p = 2$ et si $\sigma_2 = \text{Var}(X_1) < +\infty$, un phénomène surprenant (de par son universalité) se produit : la variable aléatoire

9. A condition d'ajouter une hypothèse supplémentaire sur les variables X_1, \dots, X_n)

$$\frac{1}{\sqrt{n}}(S_n - n\mathbb{E}[X_1])$$

se comporte comme une loi normale $\mathcal{N}(0; \sigma^2)$ (introduite dans la section précédente). Il s'agit du théorème de la limite centrale dont la première occurrence (dans le cas particulier de lois de Bernoulli) est dû aux mathématiciens De Moivre et Laplace; à ce propos, nous pourrions aussi nous demander ce qu'il advient lorsque le paramètre de la loi de Bernoulli impliquée dépend aussi de n , autrement dit si $p = p_n$.

Après avoir étudié des sommes des variables aléatoires, rien n'empêcherait de s'intéresser au comportement de $\max_{i=1, \dots, n} X_i$. Ce changement de fonction, entraîne rapidement des calculs plus élaborés et fournit de nouveaux résultats d'universalité.

16.4.4 Isopérimétrie et inégalités de concentration

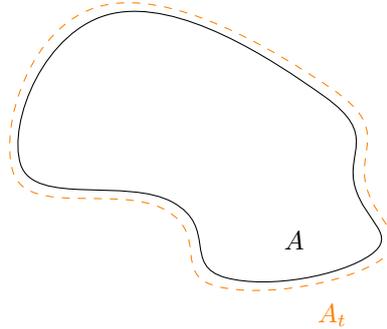
Historiquement, les inégalités de concentration (considérées dans un contexte plus large que celui exposé dans ce cours) sont liées à des problèmes isopérimétriques dont nous donnons quelques exemples ci-dessous.

Le problème isopérimétrique dans le plan euclidien est un problème concret et vieux de plusieurs millénaires. Ce problème s'énonce comme suit : quels sont les ensembles minimisant le périmètre lorsque l'aire est fixée ?

La légende raconte que la reine Didon de Carthage avait résolu, de manière pragmatique, ce problème dont les solutions sont les disques euclidiens.

Une manière d'énoncer ceci mathématiquement est la suivante : soient $A \subset \mathbb{R}^2$ et $t > 0$, nous désignons l'épaisi de taille t de cet ensemble (vis à vis de la distance euclidienne d_2) par

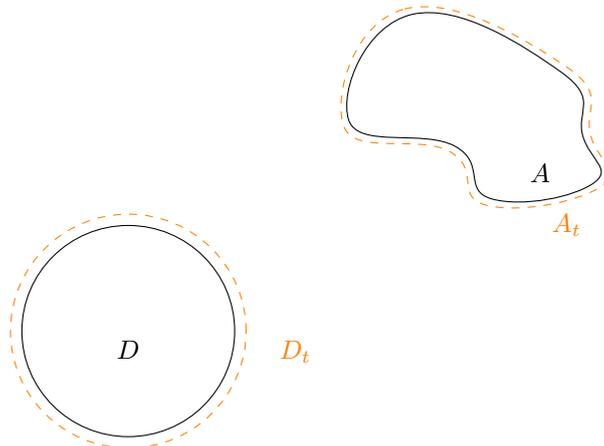
$$A_t = \{x \in \mathbb{R}^2, d_2(x, A) < t\} \quad \text{pour tout } t > 0.$$



Avec de telles notations, la solution du problème isopérimétrique euclidien s'énonce comme suit

Théorème 103 (Brunn, Minkowski, Lusternik). *Soient A un ensemble « raisonnable » du plan et D un disque tel que $\text{Aire}(A) = \text{Aire}(D)$ alors*

$$\text{Aire}(A_t) \geq \text{Aire}(D_t) \quad \text{pour tout } t > 0.$$



Autrement dit, le théorème précédent nous assure qu'en partant de deux ensembles A et D de même aires, l'aire de l'épaissi A_r sera toujours supérieur à celle de l'épaissi d'un disque.

Le problème isopérimétrique peut également s'énoncer dans un contexte probabiliste. A cet effet, considérons $X = (X_1, X_2)$ un vecteur aléatoire dont les coordonnées sont indépendantes et suivent une loi normale (i.e. $X_j \sim \mathcal{N}(0, 1)$ pour $j = 1, 2$). Etant donné un évènement A tel que $\mathbb{P}(X \in A)$ soit connu, quels sont les ensembles H (de même probabilité que A) qui minimisent la procédure d'épaississement ?

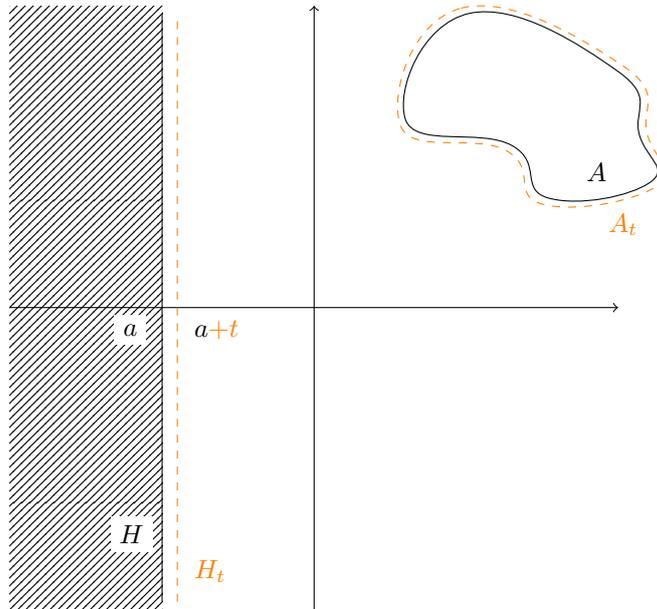
Dans ce cas, la solution du problème isopérimétrique dans ce contexte gaussien met en jeu des ensembles particuliers. Ici, les ensembles extrémaux ne sont plus des disques mais des demi-espaces H . De tels ensembles sont définis dans le plan par

$$H = \{ \vec{x} \in \mathbb{R}^2 ; \vec{x} \cdot \vec{u} \leq a \}$$

avec $\vec{u} \in \mathbb{R}^2$ un vecteur unitaire et $a \in \mathbb{R}$. La solution de ce nouveau problème isopérimétrique s'énonce alors comme suit :

Théorème 104 (Borell, Sudakov, Tsirel'son, Ibragimov). *Soit A un ensemble raisonnable du plan et H un demi-espace tel que $\mathbb{P}(X \in A) = \mathbb{P}(X \in H)$. Alors*

$$\mathbb{P}(X \in A_t) \geq \mathbb{P}(X \in H_t) \quad \text{pour tout } t \geq 0. \tag{16.4.2}$$



Remarque. La difficulté dans ce genre de résultat (la résolution d'un problème isopérimétrique) réside dans la détermination des ensembles extrémaux (les disques euclidiens ou les demi-espaces par exemple).

Nous pouvons à présent faire le lien entre problème isopérimétrique et inégalité de concentration de manière informelle. Observons tout d'abord que l'épaissi d'un ensemble complète tout l'espace dès lors que t est suffisamment grand. Ceci implique donc

$$\lim_{t \rightarrow +\infty} \mathbb{P}(X \in H_t) = \mathbb{P}(X \in \mathbb{R}^2) = 1.$$

Ou, de manière alternative, $\lim_{t \rightarrow +\infty} \mathbb{P}(X \notin H_t) = 0$. Il se trouve qu'il est possible, dans un grand nombre d'exemples, de quantifier la vitesse de décroissance vers 0 (un peu comme la décroissance en $\frac{1}{t^2}$ donnée par l'inégalité de Bienaymé-Tchebychev) et que celle-ci peut être très rapide. Ainsi, pour une certaine valeur de t (suffisamment grand), nous avons

$$\mathbb{P}(X \notin H_t) \approx 0,$$

il est donc quasiment certain que $X \in H_t$. Dans un sens, la variable X s'est concentrée dans l'ensemble H_t (par analogie avec le fait que la probabilité que l'évènement $\bar{X}_n \in]\mathbb{E}[X_1] - t; \mathbb{E}[X_1] + t[$ soit quasiment certaine d'après l'inégalité de Bienaymé-Tchebychev).

16.4.5 Constructions aléatoires

Il est parfois possible d'établir des résultats intéressants en géométrie en utilisant des constructions aléatoires.

Exemple 16.4.1. Soient $\vec{u}_1, \vec{u}_2, \vec{u}_3$ des vecteurs de l'espace de norme au plus 1. Exprimés dans une base de l'espace, nous avons

$$\vec{u}_1 = \begin{pmatrix} u_1^{(1)} \\ u_1^{(2)} \\ u_1^{(3)} \end{pmatrix} ; \quad \vec{u}_2 = \begin{pmatrix} u_2^{(1)} \\ u_2^{(2)} \\ u_2^{(3)} \end{pmatrix} ; \quad \vec{u}_3 = \begin{pmatrix} u_3^{(1)} \\ u_3^{(2)} \\ u_3^{(3)} \end{pmatrix}.$$

Considérons ensuite, $p_1, p_2, p_3 \in [0; 1]$ ainsi que le vecteur $\vec{v} = p_1 \vec{u}_1 + p_2 \vec{u}_2 + p_3 \vec{u}_3$ obtenu comme combinaison linéaire des vecteurs \vec{u}_1, \vec{u}_2 et \vec{u}_3 . Nous allons voir qu'il est possible d'approcher \vec{v} par un vecteur aléatoire $\omega \mapsto \vec{X}(\omega)$ en commettant une erreur d'au plus $\frac{\sqrt{3}}{2}$. Autrement, dit nous allons construire un vecteur aléatoire \vec{X} tel que

$$\|\vec{X} - \vec{v}\| \leq \frac{\sqrt{3}}{2}.$$

Pour cela, nous considérons des variables aléatoires indépendantes X_1, X_2 et X_3 telles que $X_i \sim B_e(p_i)$ pour $i \in \{1; 2; 3\}$ et posons

$$\vec{X} = X_1 \vec{u}_1 + X_2 \vec{u}_2 + X_3 \vec{u}_3$$

qui est un vecteur aléatoire de l'espace. L'obtention du résultat escompté va s'obtenir à la suite d'un calcul qui va combiner des propriétés vues dans le chapitre actuel (le comportement de l'espérance et de la variance vis-à-vis des sommes) mais aussi des notions impliquant des vecteurs de l'espace. En effet, nous avons

$$\begin{aligned} \mathbb{E}[\|\vec{X} - \vec{v}\|^2] &= \mathbb{E}\left[\left\|\sum_{i=1}^3 (X_i - p_i) \vec{u}_i\right\|^2\right] \\ &= \mathbb{E}\left[\sum_{k=1}^3 \sum_{i=1}^3 (X_i - p_i)^2 (u_i^{(k)})^2\right] \quad (\text{en utilisant l'indépendance et par définition de la norme } \|\cdot\|) \\ &= \mathbb{E}\left[\left(\sum_{i=1}^3 (X_i - p_i)^2\right) \times \left(\sum_{k=1}^3 (u_i^{(k)})^2\right)\right] \quad (\text{en intervertissant les sommes}) \\ &= \mathbb{E}\left[\sum_{i=1}^3 (X_i - p_i)^2 \times \|\vec{u}_i\|^2\right] \quad (\text{par définition de la norme } \|\cdot\|) \\ &= \sum_{i=1}^3 \|\vec{u}_i\|^2 \mathbb{E}[(X_i - p_i)^2] \quad (\text{par linéarité de l'espérance}) \\ &= \sum_{i=1}^3 \|\vec{u}_i\|^2 \text{Var}(X_i) \quad (\text{par définition de la variance de } X_i) \\ &= \sum_{i=1}^3 \|\vec{u}_i\|^2 p_i(1 - p_i) \quad (\text{puisque } \text{Var}(X_i) = p_i(1 - p_i)) \\ &\leq \frac{1}{4} \sum_{i=1}^3 \|\vec{u}_i\|^2 \quad (\text{puisque } p_i(1 - p_i) \leq \frac{1}{4} \text{ pour tout } i = 1, 2, 3) \\ &\leq \frac{3}{4} \quad (\text{puisque les vecteurs sont de norme au plus } 1). \end{aligned}$$

Si pour tout $\omega \in \Omega$ (Ω étant l'univers) nous avons

$$\|\vec{X}(\omega) - \vec{v}\|^2 > \frac{3}{4}$$

ceci entraîne que $\mathbb{E}\left[\|\vec{X}(\omega) - \vec{v}\|^2\right] > \frac{3}{4}$ ce qui est absurde d'après ce qui précède. Il existe donc $\omega \in \Omega$ tel que

$$\|\vec{X}(\omega) - \vec{v}\|^2 \leq \frac{3}{4}$$

d'où $\|\vec{X} - \vec{v}\| \leq \frac{\sqrt{3}}{2}$ ce qui est le résultat attendu.

Ce genre de construction aléatoire intervient également dans des problèmes de réduction de dimension.

Imaginons que nous ayons à disposition N listes composées de n éléments (avec N un nombre démesurément grand). La taille gigantesque de N ne permet pas à des algorithmes d'explorer facilement toutes ces données. Il devient alors naturel de se demander, s'il n'est pas envisageable d'oublier une partie de ces listes (i.e. considérer $k < N$ listes) sans perdre trop d'informations. Il se trouve qu'une réponse à ce problème s'obtient à l'aide de projection orthogonale dans des directions choisies aléatoirement. Ce procédé (Lemme de Johnson-Lindenstrauss) permet de se restreindre à $k = \ln N$ listes plutôt que N et la perte d'informations de cette restriction est quantifiables.

De nos jours, les méthodes de résolution impliquant de l'aléatoire deviennent de plus en plus fréquentes dans des domaines a priori éloignés des probabilités et des statistiques.

16.5 Aspects historiques

La théorie de la concentration de la mesure s'est développée durant le siècle dernier et s'est imposée comme un outil incontournable dans différents domaines des mathématiques : théorie des probabilités, statistiques, géométrie, analyse fonctionnelle

Historiquement, cette théorie est née au début des années soixante-dix lorsque Milman propose une nouvelle idée de démonstration (utilisant des inégalités isopérimétriques, lesquelles sont liées aux inégalités de concentration) du Théorème de Dvoretzky.

Durant les années 1980 et 1990, cette théorie fut révolutionnée par Talagrand et ses remarquables travaux sur le phénomène de la concentration de la mesure. Il proposa un regard novateur sur la notion d'indépendance de variables aléatoires. Il suggéra de considérer une variable aléatoire dépendant (de manière régulière) d'une grande famille de variables aléatoires comme étant quasiment constante. Cette approche lui permet d'obtenir des résultats majeurs dans cette théorie pouvant facilement s'appliquer à la résolution de problèmes complexes. Par exemple, son outil de l'inégalité de distance convexe lui a permis de développer des applications en probabilité combinatoire, en mécanique statistiques ainsi que l'étude du « maximum » de processus aléatoires. Parallèlement, une remarque de Herbst permit une approche supplémentaire, via les inégalités fonctionnelles et les équations différentielles qui y sont associées, pour étudier les propriétés de la concentration de la mesure. Il est à noter que cette approche est plus souple que l'approche isopérimétrique historique. Enfin, en 1997 une remarque de Marton, permit une troisième approche, complémentaire à celle des inégalités fonctionnelles, par le biais de la théorie du transport optimal (laquelle consiste à s'intéresser au coût minimal permettant de déplacer une quantité de matière d'un point A à un point B).

Concernant le problème isopérimétrique euclidien en dimension 2, il semblerait que sa solution, les disques euclidiens, avait déjà été démontrée de manière pragmatique par la reine Didon de Carthage. Ce résultat se généralise en dimension n et la résolution rigoureuse de ce problème n'a été trouvée qu'à la fin du 19^{ième} siècle par Brunn, Minkowski puis Lusternik.

Son analogue sphérique a été obtenu par Lévy au début des années 20 puis, par Schmidt, de manière rigoureuse dans les années 40, par des techniques délicates de symétrisation. Au milieu des années 70, le problème isopérimétrique gaussien (présenté dans le plan dans la section précédente) a été résolu par les mathématiciens Sudakov et Tsirelson, ainsi que Borell, de manière indépendante. En 1983, durant sa thèse, Ehrhard propose une démonstration plus intrinsèque et géométrique de ce résultat.

Les travaux novateurs de Milman avec ses projections aléatoires ont inspiré de nombreux domaines mathématiques : en statistique, pour des problèmes d'estimations sous certaines contraintes (méthode du LASSO par exemple) ou encore en théorie des équations aux dérivées partielles (les travaux de Burq et Lebeaux).