

Chapitre 11 – Variables aléatoires

1. Définitions

1a. Variables aléatoires

Définition : une variable aléatoire est un réel qui dépend de l'issue d'une expérience aléatoire.

Exemple 1 On propose le jeu suivant : on mise 5€, puis on lance deux fois de suite une pièce équilibrée. On gagne ensuite 4€ par Pile obtenu.

Soit X la variable aléatoire donnant le gain à ce jeu en tenant compte de la mise. Donner la loi de probabilité de X

Exemple 2

À l'issue d'une chaîne de fabrication de jouets en bois, on recherche deux types de défauts : les défauts de solidité et les défauts de couleur. Une étude a permis de relever les résultats suivants sur un échantillon de 1 000 jouets :

À réparer, un défaut de couleur coûte 5 € par jouet et un défaut de solidité coûte 12 € par jouet.

X est la variable aléatoire donnant le coût de réparation d'un objet avant d'être mis sur le marché.

- Quelles valeurs peut prendre X ?
- Décrire par une phrase l'événement $\{X \leq 10\}$.
- Que vaut $p(X = 12)$?

	Défaut de couleur	Pas de défaut de couleur	Total
Défaut de solidité	5	28	33
Pas de défaut de solidité	15	952	967
Total	20	980	1 000

Exercice 3 On considère une variable aléatoire X dont la loi de probabilité est donnée dans le tableau suivant.

- Que vaut $p(X = 8)$?
- Calculer $p(X \leq 0)$ et $p(X > 7)$.

x_i	-8	0	7	8	20
$P(X = x_i)$	0,4	0,12	0,3	...	0,08

Exercice 4 On lance deux dés équilibrés à 4 faces numérotées de 1 à 4.

On note X la variable aléatoire qui à chaque lancer associe la valeur du plus grand numéro obtenu sur les deux dés. Déterminer la loi de probabilité de X .

Exemple 1

Un arbre nous apprend que :

- la probabilité d'obtenir deux pile (donc de gagner $4 \times 2 - 5 = 3€$) est $\frac{1}{4}$,
- celle d'obtenir exactement un pile (donc de perdre 1€, car $4 - 5 = -1$) est $\frac{1}{2}$,
- celle de n'obtenir aucun pile, et donc de perdre la mise, est $\frac{1}{4}$.

On peut en déduire le tableau suivant.

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

Exemple 2

a. Le jouet peut n'avoir aucun défaut ($X = 0$), avoir seulement un défaut de solidité ($X = 12$), un défaut de couleur ($X = 5$) ou même avoir les deux défauts ($X = 17$). Ainsi, $X \in \{0; 5; 12; 17\}$.

b. L'événement correspond à « le coût de réparation est inférieur à 10€ », ce qui est équivalent à « le jouet n'a aucun défaut, ou seulement un défaut de couleur ».

c. L'événement $X = 12$ correspond aux jeux qui n'ont qu'un défaut de solidité.

$$p(X = 12) = \frac{33}{1\,000} = \mathbf{0,033}$$

Exemple 3

a. La somme des probabilités de toutes les issues est 1. Ainsi,

$$p(X = 8) = 1 - 0,4 - 0,12 - 0,3 - 0,02 = \mathbf{0,1}$$

b. $p(X \leq 0) = p(X = -8) + p(X = 0) = 0,4 + 0,12 = \mathbf{0,52}$

$p(X > 7) = p(X = 8) + p(X = 20) = 0,1 + 0,08 = \mathbf{0,18}$.

Exemple 4

Il existe $4^2 = 16$ lancers différents.

- la seule possibilité pour avoir $X = 1$ correspond au lancer où les deux dés donnent 1, donc $p(X = 1) = \frac{1}{16}$

- pour avoir $X = 2$, il faut obtenir 2 et 1, 1 et 2 ou 2 et 2. Donc $p(X = 2) = \frac{3}{16}$

- pour avoir $X = 3$, on peut compter 6 possibilités, donc $p(X = 3) = \frac{5}{16}$

- ainsi, $p(X = 4) = 1 - \frac{1}{16} - \frac{3}{16} - \frac{5}{16} = \frac{7}{16}$

x_i	1	2	3	4
$P(X = x_i)$	$\frac{\mathbf{1}}{\mathbf{16}}$	$\frac{\mathbf{3}}{\mathbf{16}}$	$\frac{\mathbf{5}}{\mathbf{16}}$	$\frac{\mathbf{7}}{\mathbf{16}}$

1b. Indépendance

Définition : deux variables aléatoires X et Y sont dites indépendantes si pour tous $x, y \in \mathbb{R}$, les événements $(X = x)$ et $(Y = y)$ sont indépendants.

Rappel : deux événements A et B de probabilité non nulle sont **indépendants** si et seulement si $p_A(B) = p(B)$ ou de façon équivalente, si $p(A \cap B) = p(A) \times p(B)$.

Par exemple, quand on lance deux pièces simultanément, les variables aléatoires X et Y donnant les résultats des deux pièces sont indépendantes.

Si A et B sont indépendants, alors leurs contraires $\bar{A}$ et $\bar{B}$ le sont aussi, de même que $\bar{A}$ et B , ou encore A et $\bar{B}$.

Exemple 1 On lance un dé équilibré à 6 faces numérotées de 1 à 6.

X est la variable aléatoire prenant la valeur 1 si le résultat est pair, et 0 sinon.

Y est la variable aléatoire prenant la valeur 1 si le résultat est supérieur ou égal à 5, et 0 sinon.

X et Y sont-elles indépendantes ?

Exemple 2 Un jeu comporte 4 cartes et sur chacune d'elle est écrit A, B, U ou W.

On tire au hasard les 4 cartes sans remise de sorte à former un mot (qui peut ne pas avoir de sens).

X est la variable aléatoire donnant le rang de la première voyelle tirée et Y est la variable aléatoire prenant la valeur 1 si on a tiré les deux voyelles en premier et 0 sinon.

Donner la loi de X et Y . Sont-elles indépendantes ?

Exemple 1 On peut chercher, par exemple, si les événements $(X = 1)$ et $(Y = 1)$ sont indépendants.

$$p(X = 1) = \frac{1}{2} \quad \text{et} \quad p(Y = 1) = \frac{2}{6} = \frac{1}{3}$$

L'événement $(X = 1 \cap Y = 1)$ correspond au cas où le résultat est pair et supérieur ou égal à 5, ce qui ne correspond qu'à face 6. Donc :

$$p(X = 1 \cap Y = 1) = \frac{1}{6}$$

De plus :

$$p(X = 1) \times p(Y = 1) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

Ainsi, les événements $(X = 1)$ et $(Y = 1)$ sont indépendants.

Les variables X et Y ne pouvant valoir que 0 ou 1, on en déduit avec la probabilités des événements contraires que les variables X et Y sont indépendantes : la valeur de l'une n'influe pas sur la valeur de l'autre.

Exemple 2

Si on tire les deux voyelles en premier (ce qui correspond à $Y = 1$), alors la première voyelle tirée est logiquement la première lettre (ce qui correspond à $X = 1$). Donc on se doute un peu que X et Y ne sont pas indépendantes.

Un tirage de mot correspond à une permutation de 4 lettres. Il y a donc $4! = 24$ mots différents.

Dans un mot de 4 lettres, il y a 2 voyelles, donc la première voyelle apparaît forcément en 3^{ème} position ou avant : X peut valoir 1, 2 ou 3.

- Il y a 12 mots dont la 1^{ère} lettre est une voyelle (on a 2 choix pour la voyelle, puis $3! = 6$ choix pour le placement des autres lettres, et $2 \times 6 = 12$).
- Il y a 4 mots où la première voyelle n'est qu'en 3^{ème} position : BWAU, BWUA, WBAU, WBUA.

On peut en déduire la loi de X .

x_i	1	2	3
$P(X = x_i)$	$\frac{12}{24} = \frac{1}{2}$	$\frac{8}{24} = \frac{1}{3}$	$\frac{4}{24} = \frac{1}{6}$

Y vaut 1 si et seulement si on a tiré les mots AUBW, AUWB, UABW ou UAWB, ce qui fait 4 mots sur 24.

y_i	0	1
$P(Y = y_i)$	$\frac{20}{24} = \frac{5}{6}$	$\frac{4}{24} = \frac{1}{6}$

Nous avons les lois de X et de Y . Montrons qu'elles ne sont pas indépendantes.

L'événement $(X = 1 \cap Y = 1)$ correspond au cas où la 1^{ère} lettre tirée est une voyelle, et où les deux voyelles ont été tirées en premier. Cela correspond aux 4 mots AUBW, AUWB, UABW ou UAWB. Ainsi :

$$p(X = 1 \cap Y = 1) = \frac{4}{24} = \frac{1}{6}$$

Or quand on effectue le produit des probabilités des deux événements séparés :

$$p(X = 1) \times p(Y = 1) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} \neq \frac{1}{6}$$

Ainsi, les variables X et Y ne sont pas indépendantes.

1c. Espérance, variance, écart-type

Définition : L'espérance $E(X)$ d'une variable aléatoire X est la **moyenne** des valeurs de X , pondérée par les probabilités.

Définitions :

- la **variance** de X est définie par :

$$V(X) = p_1(x_1 - E(X))^2 + p_2(x_2 - E(X))^2 + \dots + p_n(x_n - E(X))^2$$

- l'**écart-type** est définie par $\sigma(X) = \sqrt{V(X)}$.

Rappel :

- si X suit une loi de Bernoulli de paramètre p :

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

- si X suit une loi **binomiale** de paramètres n et p :

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

Remarques :

- Si X est une variable aléatoire dont la loi est donnée par le tableau ci-contre, l'espérance de X est $E(X) = p_1x_1 + p_2x_2 + \dots + p_nx_n$.

x_i	x_1	x_2	...	x_n
$p(X=x_i)$	p_1	p_2	...	p_n

- La variance et l'écart-type permettent de se donner une idée de la répartition des valeurs prises par une variable autour de son espérance en tenant compte des probabilités. Plus l'écart-type est grand, plus les valeurs prises par la variable aléatoire sont « éloignées » de l'espérance.

Exemple 1 On mise 3 € puis on lance trois fois de suite une pièce équilibrée. On gagne 2 € par Pile obtenu.

a. Calculer l'espérance de G , le gain obtenu. Ce jeu est-il équitable ?

b. Calculer la variance et l'écart-type de G .

x_i	-8	0	8	20	
$p(X = x_i)$	0,25	0,25	0,25	0,25	
y_i	-20	-12	-2	10	15
$p(Y = y_i)$	0,1	0,1	0,15	0,25	0,4

Exemple 2 Les lois de probabilités de deux variables aléatoires X et Y sont données ci-contre.

En utilisant la calculatrice :

a. Comparer $E(X)$ et $E(Y)$. b. Comparer $\sigma(X)$ et $\sigma(Y)$.

Exemple 1

a. On dresse un arbre si besoin, et on arrive à la loi suivante.

g_i	-5	-2	1	4
$P(G = g_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

On calcule :

$$\begin{aligned} E(G) &= -5 \times \frac{1}{8} - 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 4 \times \frac{1}{8} \\ &= -\frac{5}{8} - \frac{6}{8} + \frac{3}{8} + \frac{4}{8} = -\frac{4}{8} \\ &= -0,5 \end{aligned}$$

Le jeu n'est pas équilibré, on perd en moyenne 0,50€ en jouant.

b. On applique la formule de la variance :

$$V(G) = \frac{1}{8}(-5 - (-0,5))^2 + \frac{3}{8}(-2 - (-0,5))^2 + \frac{3}{8}(1 - (-0,5))^2 + \frac{1}{8}(4 - (-0,5))^2$$

$$V(G) = \frac{1}{8}(4,5^2 + 3 \times 1,5^2 + 3 \times 1,5^2 + 4,5^2)$$

$$V(G) = \frac{1}{8}(20,25 + 6,75 + 6,75 + 20,25)$$

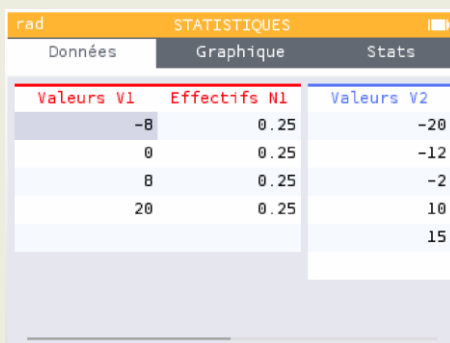
$$V(G) = \frac{54}{8}$$

$$V(G) = \mathbf{6,75}$$

$$\text{et } \sigma(G) = \sqrt{6,75} \approx \mathbf{2,60}$$

Exemple 2

Cet exemple est l'occasion de découvrir que les calculatrices savent très bien déterminer l'espérance, la variance et l'écart-type d'une série statistique. Par exemple, on peut entrer les valeurs et les « effectifs »(qui correspondent ici aux probabilités) dans le menu « Statistiques » de la Numworks.



rad STATISTIQUES		
Données		Graphique
Stats		
Valeurs V1	Effectifs N1	Valeurs V2
-8	0.25	-20
0	0.25	-12
8	0.25	-2
20	0.25	10
		15

a. On a $E(X) = 5$ et $E(Y) = 5$, les espérances sont identiques.

b. Par contre, $\sigma(X) \approx 10,34$ et $\sigma(Y) \approx 12,04$: l'écart-type est plus élevé sur la variable Y . Les valeurs de cette variable sont plus souvent éloignées de leur espérance.

2. Opérations

2a. Linéarité de l'espérance

Propriété : soient X et Y deux variables aléatoires, et $k \in \mathbb{R}$.

Alors $E(X + Y) = E(X) + E(Y)$ et $E(kX) = kE(X)$.

Exemple 1 On lance un dé à six faces numérotées de 1 à 6, et un dé à dix faces numérotés de 1 à 10.

On note X et Y les variables aléatoires donnant respectivement les résultats affichés par le dé à six faces et le dé à dix faces. Donner l'espérance de $X + Y$.

Exemple 2 Une première urne contient trois boules : deux avec le nombre 10 et une avec le nombre -3 .

Une deuxième urne contient sept boules : cinq avec le nombre 3 et deux avec le nombre 0.

On note X et Y les résultats obtenus avec chaque urne. Donner l'espérance de $X + Y$.

Exemple 1 On a

$$E(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3,5$$

$$\text{et de même, } E(Y) = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10}{10} = \frac{55}{10} = 5,5$$

On en déduit que $E(X + Y) = 3,5 + 5,5 = \mathbf{9}$.

Exemple 2

$$E(X) = \frac{2}{3} \times 10 + \frac{1}{3} \times (-3) = \frac{17}{3}$$

$$E(Y) = \frac{5}{7} \times 3 + \frac{2}{7} \times 0 = \frac{15}{7}$$

$$\text{et } E(X + Y) = \frac{17}{3} + \frac{15}{7} = \frac{119}{21} + \frac{45}{21} = \frac{164}{21} \approx \mathbf{7,81}$$

2b. Variance et indépendance

Propriété : soient X et Y deux variables aléatoires.

- si X et Y sont **indépendantes**, alors $V(X + Y) = V(X) + V(Y)$
- pour tout $k \in \mathbb{R}$, $V(kX) = k^2V(X)$ et $\sigma(kX) = |k|\sigma(X)$.

Exemple 1 On lance un dé tétraédrique équilibré dont les 4 faces portent les montants en euros 10 ; 1 ; -2 et -4. X est la variable aléatoire donnant le gain algébrique affiché.

a. Calculer $E(X)$, interpréter ce résultat et calculer $V(X)$.

b. On décide de tripler chacun des montants en jeu (par exemple 10 devient 30 ; -4 devient -12).

Z est la variable aléatoire donnant le gain algébrique à ce deuxième jeu.

Exprimer Z en fonction de X . En déduire $E(Z)$ et $V(Z)$.

Exemple 2 Une urne contient 20 000 billes dont 70 % portent le nombre 1, 25 % le nombre 2 et les autres portent le nombre 9. On effectue un prélèvement de 3 billes. On appelle S la variable aléatoire donnant la somme des trois nombres obtenus. On suppose que le prélèvement de 3 billes est assimilable à un tirage avec remise.

a. Expliquer pourquoi on peut assimiler ce prélèvement à un tirage avec remise.

b. On décompose S sous forme $S = X_1 + X_2 + X_3$ où X_i est le nombre porté par la boule lue en i -ème position. Déterminer la loi suivie par X_1 ; X_2 et X_3 , puis calculer $E(S)$ et $V(S)$.

Exemple 1 a. $E(X) = \frac{1}{4} \times 10 + \frac{1}{4} \times 1 - \frac{1}{4} \times 2 - \frac{1}{4} \times 4 = 1,25$.

En moyenne, on gagne 1,25€ en jouant à ce jeu.

$$V(X) = \frac{1}{4} (10 - 1,25)^2 + \frac{1}{4} (1 - 1,25)^2 + \frac{1}{4} (-2 - 1,25)^2 + \frac{1}{4} (-4 - 1,25)^2$$

$$V(X) = \frac{1}{4} (8,75^2 + 0,25^2 + 3,25^2 + 5,25^2)$$

$$V(X) = 28,6875$$

b. Dans ce cas, $Z = 3X$.

D'après la linéarité de l'espérance, $E(Z) = E(3X) = 3E(X) = 3 \times 1,25 = 3,75$ En moyenne, on gagne 3,75€ en jouant à ce nouveau jeu.

$$V(Z) = V(3X) = 3^2V(X) = 9V(X) = 9 \times 28,6875 = 258,1875.$$

Exemple 2 a. Le nombre de billes est suffisamment grand : tirer 3 billes sur 20 000 ou sur 19 997 billes ne change presque pas les probabilités.

b. • X_1, X_2 et X_3 suivent la même loi :

x_i	1	2	3
$P(X_1 = x_i)$	0,7	0,25	0,05

$$\bullet E(X_1) = 1 \times 0,7 + 2 \times 0,25 + 3 \times 0,05 = 0,7 + 0,5 + 0,15 = 1,35$$

Comme X_2 et X_3 suivent la même loi :

$$E(S) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 3 \times E(X_1) = 3,95$$

$$\bullet \text{ On calcule la variance de } X_1 \text{ à la calculatrice. } V(X_1) = 0,3275$$

Le tirage est assimilé à un tirage avec remise, donc on peut considérer que **les variables X_1, X_2 et X_3 sont indépendantes.**

Il est indispensable de préciser cela pour calculer la somme des variances.

$$\text{Ainsi, } V(S) = V(X_1 + X_2 + X_3) = V(X_1) + V(X_2) + V(X_3) = 3 \times V(X_1) = 0,9825.$$

2c. Échantillons

Définition : soient $X_1, X_2, \dots, X_n$ plusieurs variables aléatoires indépendantes, qui suivent toutes la même loi.

Le n -uplet $(X_1, X_2, \dots, X_n)$ est appelé **échantillon de taille n** associé à cette loi.

Propriété : avec les notations précédentes, on a :

$$E(X_1 + \dots + X_n) = nE(X_1) \quad V(X_1 + \dots + X_n) = nV(X_1)$$

Remarques • Si on lance 10 dés à six faces, soit $X_1, X_2, \dots, X_{10}$ les variables représentant les résultats. Le 10-uplet $(X_1; X_2; \dots; X_{10})$ est un échantillon de taille 10 associé à la loi d'un lancer de dé à six faces. L'échantillon peut prendre 6^{10} valeurs différentes. Par exemple : $(1; 5; 5; 4; 2; 6; 6; 3; 5; 2)$ est une de ces valeurs.

• Si $X_1, X_2, \dots, X_n$ sont n variables aléatoires indépendantes qui suivent une **loi de Bernoulli de paramètre p** (elles valent 1 en cas de succès avec probabilité p , et 0 en cas d'échec avec probabilité $1 - p$).

Le n -uplet $(X_1; X_2; \dots; X_n)$ est un échantillon de taille n , associé à cette loi. Les valeurs possibles pour cet échantillon peuvent se représenter sur un arbre, il y en a 2^n . Par exemple : $(1; 1; 0; 1; 0; \dots)$

La loi de la **somme** $S = X_1 + X_2 + \dots + X_n$ est la **loi binomiale $\mathcal{B}(n; p)$** .

Exemple 1 Dans un pays, 27% de la population compte voter pour un candidat A à la prochaine élection. On interroge 40 personnes de ce pays. Pour tout entier i entre 1 et 40, X_i est la variable aléatoire valant 1 si la i -ème personne interrogée compte voter pour A, et 0 sinon.

À quoi correspond la somme $S = X_1 + X_2 + \dots + X_{40}$? Quelle est la loi suivie par S ? Calculer $E(S)$ et $V(S)$.

Exemple 2 On mise 2€ puis on lance un dé équilibré à quatre faces.

On gagne 1€ si on fait un 3, et 10€ si on fait un 4. Sinon on ne gagne rien.

On joue 15 fois à ce jeu. Z est la variable aléatoire donnant le gain algébrique total.

Calculer $E(Z)$ et en donner une interprétation. Puis calculer $V(Z)$.

Exemple 3 $Y_1, Y_2, \dots, Y_{30}$ sont des variables aléatoires indépendantes suivant une même loi binomiale de paramètres $n = 12$ et $p = 0,48$. On pose $S = Y_1 + Y_2 + \dots + Y_{30}$

Calculer $E(S)$ et $V(S)$. Que remarque-t-on ?

Exemple 1 La somme S correspond au nombre de personnes qui comptent voter pour le candidat A sur les 40 interrogées.

Chaque X_i suit une loi de Bernoulli de paramètre 0,27 et les X_i sont indépendantes, donc S suit la loi binomiale notée **$\mathcal{B}(40; 0,27)$** .

On a donc $E(S) = 40 \times 0,27 = 10,8$ et $V(S) = 40 \times 0,27 \times 0,73 = 7,884$.

Exemple 2

Il s'agit d'un échantillon : tous les X_i sont indépendantes et suivent la même loi.

Prenons par exemple X_1 qui suit la loi suivante :

x_i	-2	-1	8
$P(X_1 = x_i)$	0,5	0,25	0,25

• On a alors $E(X_1) = -2 \times 0,5 - 1 \times 0,25 + 8 \times 0,25 = 0,75$.

Ainsi, $E(Z) = E(X_1 + X_2 + \dots + X_{15}) = 15 \times E(X_1) = 15 \times 0,75 = \mathbf{11,25}$.

Cela signifie qu'en moyenne, en jouant 15 fois à ce jeu, on gagne 11,25€.

• $V(X_i) = 0,5(-2 - 0,75)^2 + 0,25(1 - 0,75)^2 + 0,25(8 - 0,75)^2 = 17,6875$

Donc par indépendance des X_i , $V(Z) = 15 \times V(X_1) = \mathbf{265,3125}$

Exemple 3

Comme les Y_i forment un échantillon de taille 30, on a :

$E(S) = 30 \times E(Y_1) = 30 \times 12 \times 0,48 = 360 \times 0,48 = \mathbf{172,8}$

$V(S) = 30 \times V(Y_1) = 30 \times 12 \times 0,48 \times 0,52 = 360 \times 0,48 \times 0,52 = \mathbf{89,856}$

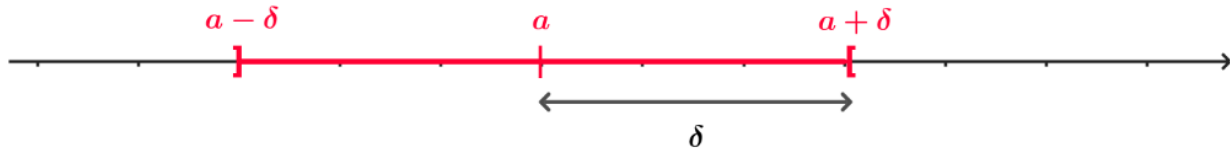
On remarque que cela correspond à l'espérance et la variance d'une variable de loi $\mathcal{B}(360; 0,48)$.

3. Concentration

3a. Intervalles centrés

Rappel : soit $a \in \mathbb{R}$, et soit $\delta > 0$.

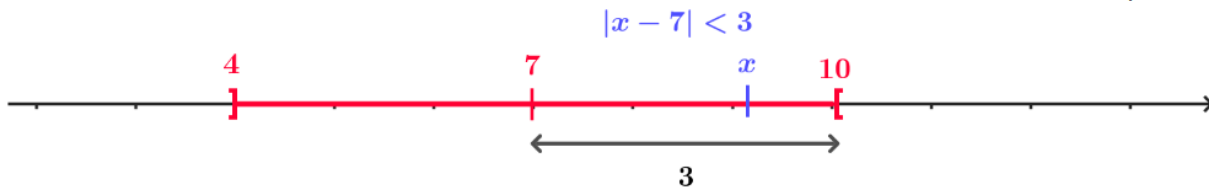
L'intervalle $]a - \delta; a + \delta[$ est un **intervalle centré en a** , et contient tous les nombres à une distance de a strictement inférieure à δ .



Si un réel x appartient à cet intervalle, on peut utiliser la notation valeur absolue : $|x - a| < \delta$.

Exemple : si $a = 7$ et $\delta = 3$, il s'agit de l'intervalle $]4; 10[$.

Tous les réels x de cet intervalle sont à une distance de 7 strictement inférieure à 3. On écrit $|x - 7| < 3$.



3b. Inégalité de Bienaymé-Tchebychev

Soit X une v.a. On note $\mu = E(X)$ son espérance et V sa variance.
On s'intéresse à la probabilité que X s'écarte de son espérance μ .

Propriété : Pour tout écart $\delta > 0$, on a :

$$p(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$$

$$\text{c'est-à-dire } p(X \notin]\mu - \delta; \mu + \delta[) \leq \frac{V}{\delta^2}$$

Avec les probabilités des événements contraires, on obtient :

$$p(|X - \mu| < \delta) \geq 1 - \frac{V}{\delta^2}$$

$$\text{c'est-à-dire } p(X \in]\mu - \delta; \mu + \delta[) \geq 1 - \frac{V}{\delta^2}$$

Dans l'inégalité de Bienaymé-Tchebychev, le δ représente l'écart entre X et son espérance $E(X)$.

Par exemple, si $E(X) = 8$ et $\delta = 5$, alors la formule nous dit que $p(|X - 8| \geq 5) \leq \frac{V(X)}{5^2}$.

Autrement dit, X a une probabilité inférieure à $\frac{V(X)}{5^2}$ de ne PAS être dans l'intervalle $]3; 13[$.

Il est important de toujours traduire les probabilités demandées par des **intervalles ouverts**.

Exemple 1 Dans un cabinet médical, le nombre de patients vus chaque jour par un médecin est donné par une variable aléatoire N d'espérance $E(N) = 32$ et d'écart type $\sigma(N) = 3$.

a. En appliquant l'inégalité de Bienaymé-Tchebychev, que peut-on dire de $p(|N - 32| \geq 6)$?

Interpréter le résultat dans les termes de l'énoncé.

b. Donner une minoration de la probabilité que ce médecin voie entre 22 et 42 patients dans une journée.

Exemple 2 Une société livre des téléviseurs. Lorsqu'une commande est passée par un client, on considère que le temps de livraison est modélisé par une variable aléatoire T égale à la somme de deux variables T_1 et T_2 .

La variable aléatoire T_1 modélise le nombre entier de jours pour l'acheminement du téléviseur depuis un entrepôt de stockage vers une plateforme de distribution. La variable aléatoire T_2 modélise le nombre entier de jours pour l'acheminement du téléviseur depuis cette plateforme jusqu'au domicile du client.

On admet que les variables aléatoires T_1 et T_2 sont indépendantes, et on donne :

• L'espérance $E(T_1) = 4$ et la variance $V(T_1) = 2$. • L'espérance $E(T_2) = 3$ et la variance $V(T_2) = 1$

a. Déterminer l'espérance $E(T)$ et la variance $V(T)$ de la variable aléatoire T .

b. Un client passe une commande de téléviseur sur internet. Justifier que la probabilité qu'il reçoive son téléviseur entre 5 et 9 jours après sa commande est supérieure ou égale à $\frac{2}{3}$.

Exemple 3 Une société demande à un institut de sondage de faire une enquête sur le profil de ses clients réguliers. L'institut a élaboré un questionnaire en ligne constitué d'un nombre variable de questions. On choisit au hasard un échantillon de 1 000 clients réguliers, à qui le questionnaire est proposé. On considère que ces 1 000 clients répondent.

- Pour les remercier, la société offre un bon d'achat à chacun des clients de l'échantillon.

Le montant de ce bon d'achat dépend du nombre de questions posées au client.

- La société souhaite récompenser particulièrement les clients qui ont acheté une carte de fidélité et, en plus du bon d'achat, offre à chacun d'eux une prime d'un montant de 50 euros versée sur la carte de fidélité.

On note Y_1 la variable aléatoire qui, à chaque échantillon de 1 000 clients réguliers, associe le total, en euros, des montants du bon d'achat des 1 000 clients. On admet que $E(Y_1) = 30\,000$ et $V(Y_1) = 100\,000$.

On note X_2 la variable aléatoire qui, à chaque échantillon de 1 000 clients réguliers, associe le nombre de clients ayant acheté la carte de fidélité parmi eux, et on note Y_2 la variable aléatoire qui, à chaque échantillon de 1 000 clients, associe le total, en euros, des montants de la prime de fidélité versée.

On admet que X_2 suit la loi binomiale de paramètres 1 000 et 0,47.

1. Expliquer pourquoi $Y_2 = 50X_2$.

2. Calculer l'espérance $E(X_2)$ de la variable X_2 et interpréter le résultat dans le contexte de l'exercice.

On note $Y = Y_1 + Y_2$ la variable aléatoire égale au total général, en euros, des montants offerts (bon d'achat et prime de fidélité) aux 1 000 clients. On admet que les variables aléatoires Y_1 et Y_2 sont indépendantes.

3. On note Z la variable aléatoire définie par $Z = \frac{Y}{1\,000}$. Préciser ce que modélise la variable Z dans le contexte de l'exercice. Vérifier que son espérance $E(Z)$ est égale à 53,5 et que sa variance $V(Z)$ est égale à 0,722 75.

4. À l'aide de l'inégalité de Bienaymé-Tchebychev, vérifier que la probabilité que Z soit strictement compris entre 51,7 euros et 55,3 euros est supérieure à 0,75.

*Comme précisé avant les exemples, il faut bien remarquer qu'on doit travailler avec des **intervalles ouverts**, dont les bornes sont **exclues**.*

Exemple 1

a. Tout d'abord, l'écart-type étant la racine carrée de la variance,

$$V(N) = \sigma(N)^2 = 3^2 = 9.$$

D'après l'inégalité de Bienaymé-Tchebychev :

$$p(|N - 32| \geq 6) \leq \frac{9}{6^2} \Leftrightarrow p(|N - 32| \geq 6) \leq \frac{1}{4}$$

La probabilité que N s'écarte d'au moins 6 patients de sa moyenne (c'est-à-dire, la probabilité que N ne soit PAS compris entre 26 et 38 exclus) est inférieure à $\frac{1}{4}$.

En d'autres termes, cela signifie que la probabilité que le nombre de patients soit compris entre 26 et 38 exclus (donc entre 27 et 37) est supérieure à $\frac{3}{4}$.

b. On cherche la probabilité que $N \in [22; 42]$, c'est-à-dire $N \in]21; 43[$.

On applique donc Bienaymé-Tchebychev avec $\delta = 11$.

$$p(|N - 32| < 11) \geq 1 - \frac{9}{11^2} \Leftrightarrow p(|N - 32| < 11) \geq \frac{112}{121}$$

Le médecin a donc au moins environ **93%** de probabilité de voir entre 22 et 42 patients.

Exemple 2

a. Par linéarité de l'indépendance, $E(T) = E(T_1) + E(T_2) = 4 + 3 = 7$.

Par indépendance des variables T_1 et T_2 , $V(T) = V(T_1) + V(T_2) = 2 + 1 = 3$.

b. On cherche la probabilité que $T \in [5; 9]$, c'est-à-dire $T \in]4; 10[$.

On applique Bienaymé-Tchebychev avec $\delta = 3$.

$$p(|T - 7| < 3) \geq 1 - \frac{3}{3^2} \Leftrightarrow p(|T - 7| < 3) \geq \frac{2}{3}$$

Exemple 3 *Tout cet énoncé indigeste pour 4 pauvres questions, mais c'est le Bac.*

1. X_2 est le nombre de clients ayant une carte de fidélité, à qui on a versé 50€ chacun, et Y_2 correspond au **montant total versé à ces clients**, donc $Y_2 = 50X_2$.

2. X_2 suit une loi binomiale $\mathcal{B}(1\,000; 0,47)$, donc $E(X_2) = 1\,000 \times 0,47 = 470$.

Cela correspond au nombre moyen de clients qui ont la carte de fidélité.

3. *Il faut utiliser soigneusement les propriétés vues en parties 2a et 2b du cours.*

• Y est le total des montants versés aux clients, donc $Z = \frac{Y}{1\,000}$ est le **montant moyen** versé à un client.

• Par linéarité de l'espérance :

$$E(Z) = \frac{E(Y)}{1\,000} = \frac{E(Y_1) + E(Y_2)}{1\,000}$$

Or $Y_2 = 50X_2$, donc $E(Y_2) = 50 \times E(X_2) = 50 \times 470 = 23\,500$. Ainsi :

$$E(Z) = \frac{30\,000 + 23\,500}{1\,000} = \frac{53\,500}{1\,000} = 53,5$$

• Passons maintenant à la variance.

$$V(Z) = V\left(\frac{Y}{1\,000}\right) = \left(\frac{1}{1\,000}\right)^2 \times V(Y) = \frac{V(Y)}{1\,000\,000}$$

Or Y_1 et Y_2 sont indépendantes, donc $V(Y) = V(Y_1) + V(Y_2)$.

$V(Y_1) = 100\,000$ est donnée dans l'énoncé, et

$$V(Y_2) = V(50X_2) = 50^2 \times V(X_2) = 2\,500 \times V(X_2)$$

Or X_2 suit une loi $\mathcal{B}(1\,000; 0,47)$, donc $V(X_2) = 1\,000 \times 0,47 \times 0,53 = 249,1$.

Ainsi, $V(Y_2) = 2\,500 \times 249,1 = 622\,750$. Enfin,

$$V(Z) = \frac{V(Y)}{1\,000\,000} = \frac{V(Y_1) + V(Y_2)}{1\,000\,000} = \frac{100\,000 + 622\,750}{1\,000\,000} = 0,722\,75.$$

4. On cherche la probabilité que $Z \in]51,7; 55,3[$, qui est un intervalle ouvert centré sur l'espérance $E(Z) = 53,5$ avec $\delta = 1,8$.

$$p(|Z - 53,5| < 1,8) \geq 1 - \frac{0,722\,75}{1,8^2}$$

Or $1 - \frac{0,722\,75}{1,8^2} \approx 0,78$ donc la probabilité que $Z \in]51,7; 55,3[$ est **bien supérieure à 0,75**.

3c. Inégalité de concentration

Propriété : Soit $(X_1; X_2; \dots; X_n)$ un échantillon de variables aléatoires d'espérance μ et de variance V , et $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ la moyenne de cet échantillon.

Pour tout réel strictement positif δ , on a l'inégalité :

$$p(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$$
$$\text{et } p(|M_n - \mu| < \delta) \geq 1 - \frac{V}{n\delta^2}$$

Exemple 1

a. Avant une élection, 20% d'une population compte voter pour un candidat A.

On sonde 1000 personnes parmi la population générale. Quelle est la probabilité que le résultat du sondage s'écarte de plus de 3% de la proportion de votants pour le candidat A ?

b. Même question avec le candidat B, pour lequel 5% de la population compte voter.

Exemple 2 Après un examen, on interroge au hasard dix étudiants.

Les variables aléatoires $N_1, N_2, \dots, N_{10}$ modélisent la note sur 20 obtenue à l'examen par chacun d'entre eux.

On admet que ces variables sont indépendantes et suivent la même loi binomiale de paramètres $(20; 0,615)$.

Soit S la variable définie par $S = N_1 + N_2 + \dots + N_{10}$.

1. Calculer l'espérance $E(S)$ et la variance $V(S)$ de la variable aléatoire S .

2. On considère la variable aléatoire $M = \frac{S}{10}$. Que modélise cette variable M dans le contexte de l'exercice ?

3. Justifier que $E(M) = 12,3$ et $V(M) = 0,47355$.

4. Justifier que la probabilité que la moyenne des notes de dix étudiants pris au hasard soit strictement comprise entre 10,3 et 14,3 est d'au moins 80 %.

L'inégalité de concentration est juste un raccourci pour Bienaymé-Tchebychev dans le cas où on a un échantillon. On n'est pas obligés de s'en servir, comme on le verra en exemple 2.

Exemple 1

a. On considère X_i , variable égale à 1 si la i -ème personne interrogée compte voter pour A, et 0 sinon. X_i suit la loi de Bernoulli de paramètre 0,2.

Son espérance est 0,2 et sa variance est $V = 0,2 \times 0,8 = 0,16$.

$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ correspond alors au résultat du sondage : c'est la proportion de personnes interrogées qui compte voter pour A.

$$p(|M_n - 0,2| \geq 0,03) \leq \frac{0,16}{1000 \times 0,03^2} \Leftrightarrow p(|M_n - 0,2| \geq 0,03) \Leftrightarrow \frac{0,16}{0,9}$$

ce dernier résultat étant à peu près égal à 18%.

Cela signifie en particulier que le résultat du sondage a 82% de probabilité d'être dans l'intervalle]17% ; 23%[, ce qui est acceptable. Les « vrais » sondages font mieux.

b. Ici, les X_i suivent une loi de Bernoulli de paramètre 0,05, donc $V = 0,05 \times 0,95 = 0,0475$.

$$p(|M_n - 0,05| \geq 0,03) \leq \frac{0,0475}{1000 \times 0,03^2} \Leftrightarrow p(|M_n - 0,05| \geq 0,03) \Leftrightarrow \frac{0,0475}{0,9}$$

ce qui fait **environ 5%**.

Le résultat du sondage a 95% de probabilité d'être dans l'intervalle]2% ; 8%[, encore heureux.

Exemple 2

1. On a $E(N_1) = 20 \times 0,615 = 12,3$ et $V(N_1) = 20 \times 0,615 \times 0,385 = 4,7355$. Ainsi, $E(S) = 10 \times E(N_1) = 123$ et $V(S) = 10 \times V(N_1) = 47,355$.

2. S représente la somme des 10 notes, donc M correspond à la **moyenne des 10 notes**.

3. D'après la linéarité de l'espérance :

$$E(M) = E\left(\frac{S}{10}\right) = \frac{E(S)}{10} = \frac{123}{10} = 12,3$$

D'après la règle de calcul sur la variance :

$$V(M) = V\left(\frac{S}{10}\right) = \frac{V(S)}{10^2} = \frac{47,355}{100} = 0,47355$$

4. L'intervalle]10,3; 14,3[est centré sur 12,3 avec $\delta = 2$.

Attention : on peut soit appliquer Bienaymé-Tchebychev avec la variable M , dont on connaît maintenant l'espérance et la variance, soit appliquer l'inégalité de concentration sur les N_i .

• Avec Bienaymé-Tchebychev sur la variable M :

$$p(|M - 12,3| < 2) \leq 1 - \frac{0,47355}{2^2}$$

ce dernier résultat valant environ 0,88, ce qui résout la question.

• Avec l'inégalité de concentration sur les N_i , d'espérance 12,3 et de variance 4,7355 :

$$p(|M - 12,3| < 2) \leq 1 - \frac{4,7355}{10 \times 2^2}$$

et ce résultat correspond presque au même calcul et vaut aussi 0,88.

4. Loi faible des grands nombres

Soit $(X_1, X_2, \dots, X_n)$ un échantillon. Soit $M_n = \frac{X_1 + \dots + X_n}{n}$ sa moyenne.

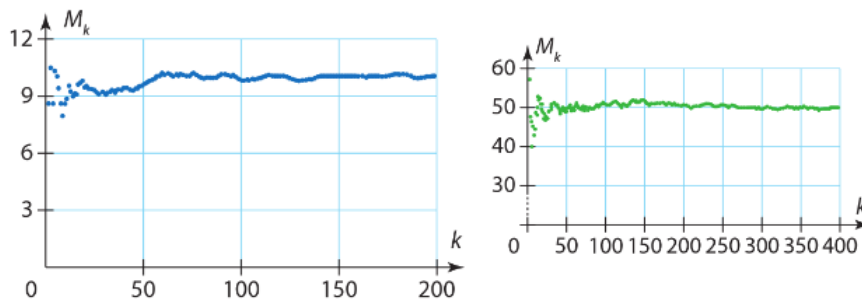
Pour tout $\delta > 0$, on a :

$$\lim_{n \rightarrow +\infty} p(|M_n - \mu| \geq \delta) = 0$$

La loi des grands nombres dit que lorsque le nombre de tirages tend vers $+\infty$, la probabilité que l'échantillon s'éloigne de sa moyenne tend vers 0.

Remarque : en revanche, on ne sait rien sur la vitesse de convergence : la loi des grands nombres ne dit pas « à quelle vitesse » la moyenne de l'échantillon se rapprochera de l'espérance attendue.

Exemples : dans chaque cas, on considère un échantillon $(X_1; X_2; \dots; X_n)$ de variables aléatoires d'espérance μ , et on représente la moyenne M_k des k premiers échantillons. Estimer μ .



Attention : cette loi ne permet pas de prévoir le résultat d'expériences.

Par exemple, si on lance 9 pièces et qu'on obtient « pile », la loi des grands nombres ne dit pas que le 10^{ème} lancer donnera « face » pour rééquilibrer. Elle dit juste qu'avec un grand nombre de lancers, il devient de plus en plus improbable de s'écarter de la moyenne attendue.

Ainsi, la probabilité d'obtenir 9 « pile », comme supposé précédemment, n'est même pas de 1%.

Dans le premier cas, $\mu \approx 10$, et dans le deuxième cas, $\mu \approx 50$.