

CHAPITRE 10 : Loi des grands nombres

1	Transformation affine d'une variable aléatoire	2
1.1	Activité.....	2
1.2	Transformation affine d'une variable aléatoire X	5
2	Somme de deux variables aléatoires.....	5
2.1	Activité.....	5
2.2	Somme de deux variables aléatoires.....	9
3	Démonstration des formules de l'espérance et de la variance de la loi binomiale	9
3.1	Décomposition d'une variable aléatoire X suivant une loi binomiale $\mathbf{B}(n, p)$	10
3.2	Épreuve de Bernoulli, variable X_1 suivant une loi de Bernoulli	10
3.3	Formules de l'espérance et de la variance pour une variable aléatoire X qui suit une loi binomiale.....	11
4	Échantillon de taille n d'une loi de probabilité X.....	16
4.1	Définition d'un échantillon de taille n d'une loi de probabilité sur un exemple.....	16
4.2	Espérance, variance et écart-type de la somme $S_n = X_1 + X_2 + \dots + X_n$ et de la moyenne empirique $M_n = S_n/n$ à l'aide de l'exemple précédent	16
4.3	Définition d'un échantillon d'une loi de probabilité	19
5	Inégalité de Bienaymé-Tchebychev.....	21
5.1	Inégalité de Bienaymé-Tchebychev sur un exemple.....	21
5.2	Formule de l'inégalité de Bienaymé-Tchebychev.....	24
5.3	Exemple d'utilisation de l'inégalité de Bienaymé-Tchebychev	26
5.4	Cas de la loi binomiale.....	26
6	Inégalité de concentration de la moyenne empirique.....	30
7	Loi des grands nombres.....	34

CHAPITRE 10 : Loi des grands nombres

1 Transformation affine d'une variable aléatoire

1.1 Activité

On considère le jeu suivant :

On tourne une roue coupée en 10 sections de même aire.

5 sections portent la mention « -1 » signifiant que l'on perd 1 euro si la roue s'arrête sur l'une de ces sections.

2 sections portent la mention « 0 » signifiant que l'on ne gagne rien.

2 sections portent la mention « 1 » signifiant que l'on gagne un euro.

1 section porte la mention « 2 » signifiant que l'on gagne 2 euros.

On note X la variable aléatoire égale au gain algébrique (pouvant être négatif en cas de perte) du joueur lors d'une partie.

1) a) Donner la loi de probabilité de la variable aléatoire X .

Réponse

x_i	-1	0	1	2	TOTAL
p_i	$\frac{5}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	1

b) Calculer l'espérance $E(X)$ la variance $V(X)$ et l'écart-type $\sigma(X)$.

Réponse

$$E(X) = \sum_{i=1}^4 p_i x_i$$

$$E(X) = \frac{5}{10} \times (-1) + \dots + \frac{1}{10} \times (2) \quad \underline{E(X) = -0,1}$$

$$V(X) = \sum_{i=1}^4 p_i x_i^2 - (E(X))^2$$

$$V(X) = \frac{5}{10} \times (-1)^2 + \dots + \frac{1}{10} \times (2)^2 - (0,1)^2 \quad \underline{V(X) = 1,09}$$

$$\sigma(X) = \sqrt{1,09} \quad \underline{\sigma(X) \approx 1,044}$$

2) On décide de garder le même principe que précédemment mais de modifier les gains et pertes mentionnés sur les sections de la manière suivante : On double les valeurs puis on leur ajoute 3.

On note Y la variable aléatoire égale au gain algébrique du joueur lors d'une partie avec cette nouvelle situation.

a) Exprimer la variable aléatoire Y en fonction de la variable aléatoire X .

Réponse

$$Y = 2X + 3$$

b) Donner la loi de probabilité de la variable aléatoire Y .

Réponse

y_i	1	3	5	7	TOTAL
p_i	$\frac{5}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	1

c) Calculer $E(Y)$. Comparer $E(Y)$ à $2E(X) + 3$.

Réponse

$$E(Y) = \sum_{i=1}^4 p_i y_i$$

$$E(Y) = \frac{5}{10} \times (1) + \dots + \frac{1}{10} \times (7)$$

$$\underline{E(Y) = 2,8}$$

D'autre part :

$$2E(X) + 3 = 2 \times -0,1 + 3$$

$$\underline{2E(X) + 3 = 2,8}$$

Donc $E(Y) = 2E(X) + 3$

c'est-à-dire

$$\mathbf{E(2X + 3) = 2E(X) + 3}$$

d) Calculer $V(Y)$. Comparer $V(Y)$ à $2^2 V(X)$.

Réponse

$$V(Y) = \sum_{i=1}^4 p_i y_i^2 - (E(Y))^2$$

$$V(Y) = \frac{5}{10} \times (1)^2 + \dots + \frac{1}{10} \times (7)^2 - (2,8)^2$$

$$\underline{V(Y) = 4,36}$$

D'autre part :

$$\underline{2^2 V(X) = 4 \times 1,09 = 4,36}$$

$$\text{Donc } V(Y) = 2^2 V(X)$$

c'est-à-dire

$$\mathbf{V(2X + 3) = 2^2 V(X)}$$

e) Calculer $\sigma(Y)$. Comparer $\sigma(Y)$ à $|2|\sigma(X)$.

Réponse

$$\sigma(Y) = \sqrt{V(Y)}$$

$$\sigma(Y) = \sqrt{4,36}$$

D'autre part :

$$|2|\sigma(X) = 2\sqrt{1,09}$$

$$|2|\sigma(X) = \sqrt{4(1,09)}$$

$$|2|\sigma(X) = \sqrt{4,36}$$

$$\text{Donc } \sigma(Y) = |2|\sigma(X)$$

c'est-à-dire

$$\mathbf{\sigma(2X + 3) = |2|\sigma(X)}$$

1.2 Transformation affine d'une variable aléatoire X

Soit X une variable aléatoire et a et b deux nombres réels.

Les valeurs prises par X sont x_1, x_2, \dots, x_n .

La variable aléatoire Y définie par $Y = aX + b$ est la variable aléatoire qui prend pour valeurs les réels :

$$y_1 = ax_1 + b, y_2 = ax_2 + b, \dots, y_n = ax_n + b.$$

La variable Y est obtenue à partir de la variable X par une transformation affine (par analogie avec l'expression d'une fonction affine $f(x) = ax + b$).

Si $Y = aX + b$ alors on a :

$$E(Y) = E(aX + b) = aE(X) + b$$

La multiplication par a et l'addition de b ont toutes les deux une influence sur l'espérance.

$$V(Y) = V(aX + b) = V(aX) = a^2V(X).$$

On remarque que, puisque $V(aX + b) = V(aX)$, l'addition d'une constante b à toutes les valeurs de X n'a pas d'influence sur la variance c'est-à-dire sur la dispersion de X .

$$\sigma(Y) = |a| \sigma(X).$$

2 Somme de deux variables aléatoires

2.1 Activité

On considère le jeu suivant qui se déroule en deux parties. Ces deux parties sont obligatoires. On ne peut pas s'arrêter après la première partie.

Première partie :

- On lance un dé équilibré à 6 faces numérotées de 1 à 6. On regarde si on tombe sur un nombre pair ou impair. Si le nombre est pair alors on gagne 10 €. Si le nombre est impair on gagne 5 €.

Deuxième partie :

On relance le dé. Si on tombe sur 1 ou 2 alors on gagne 5 €. Si on tombe sur 3 alors on gagne 2 €. Si on tombe sur 4, 5 ou 6 on ne gagne rien.

X est la variable aléatoire donnant le gain à la première partie. Y est la variable aléatoire donnant le gain à la deuxième partie.

1) Donner les lois de probabilités de X et Y .

Réponse

x_i	5	10	TOTAL
p_i	$\frac{3}{6}$	$\frac{3}{6}$	1

y_i	0	2	5	TOTAL
p_i	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	1

2) Calculer $E(X)$ et $E(Y)$ ainsi que $V(X)$ et $V(Y)$.

Réponse

$$E(X) = \sum_{i=1}^2 p_i x_i$$

$$E(X) = \frac{3}{6} \times (5) + \frac{3}{6} \times (10) \quad \underline{E(X) = 7,5 \text{ €}}$$

$$E(Y) = \sum_{i=1}^3 p_i y_i$$

$$E(Y) = \frac{3}{6} \times (0) + \frac{1}{6} \times (2) + \frac{2}{6} \times (5) \quad \underline{E(Y) = 2 \text{ €}}$$

$$V(X) = \sum_{i=1}^2 p_i x_i^2 - (E(X))^2$$

$$V(X) = \frac{3}{6} \times (5)^2 + \frac{3}{6} \times (10)^2 - (7,5)^2 \quad \underline{V(X) = 6,25}$$

$$V(Y) = \sum_{i=1}^3 p_i y_i^2 - (E(Y))^2$$

$$V(Y) = \frac{3}{6} \times (0)^2 + \frac{1}{6} \times (2)^2 + \frac{2}{6} \times (5)^2 - (2)^2 \quad \underline{V(Y) = 5}$$

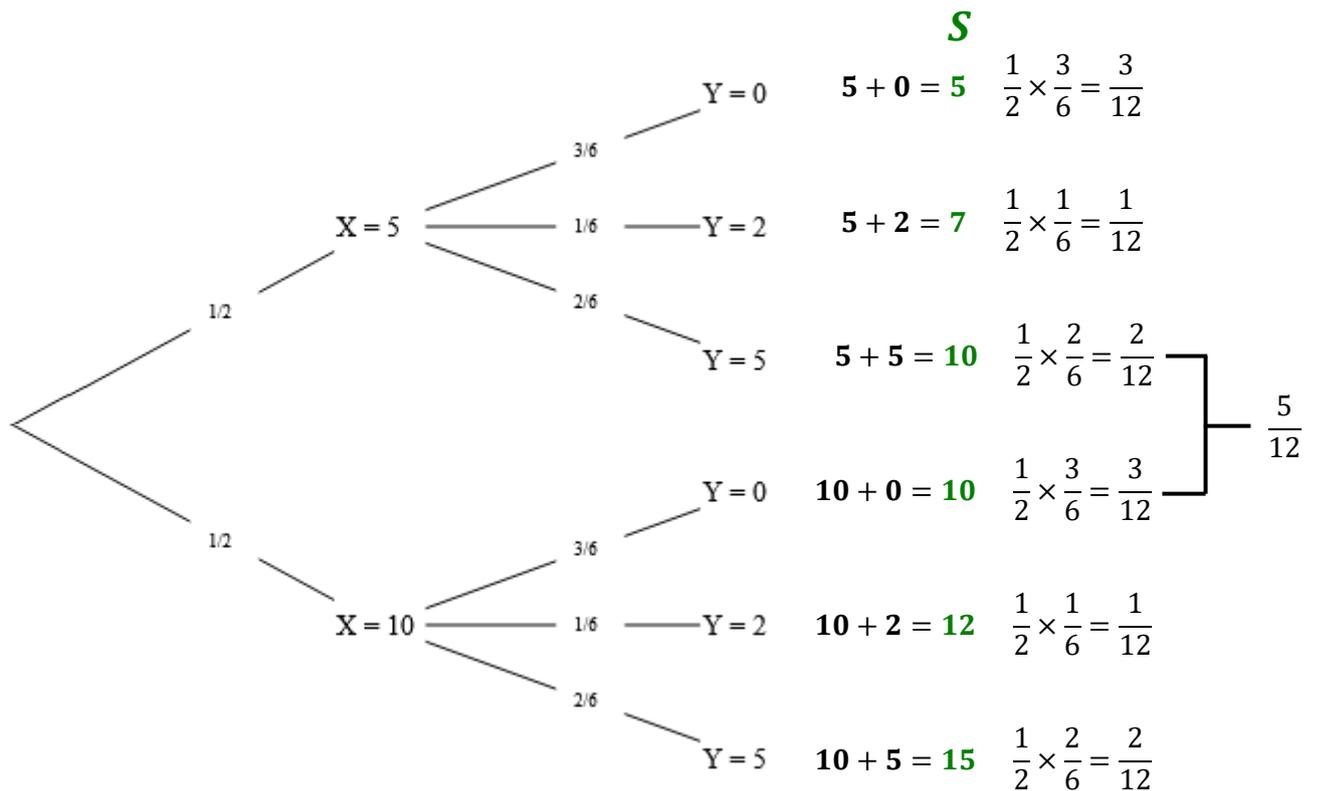
3) On note S la variable aléatoire donnant la somme des gains des deux jeux.

a) Exprimer S en fonction des variables aléatoires X et Y .

Réponse : $S = X + Y$

b) Donner la loi de probabilité de la variable aléatoire S qui donne le gain total cumulé.

Réponse : On fait un arbre pour visualiser toutes les possibilités et en regrouper certaines.



On résume dans un tableau la somme de tous les gains possibles.

Sur la première ligne on a mis toutes les valeurs possibles prises par X .

Sur la première colonne on a mis toutes les valeurs possibles prises par Y .

$Y \backslash X$	5	10
0	5	10
2	7	12
5	10	15

On en déduit la loi de probabilité de la variable aléatoire somme S

s_i	5	7	10	12	15	TOTAL
p_i	$\frac{3}{12}$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	1

c) Calculer $E(S)$.

Réponse

$$E(S) = \sum_{i=1}^5 p_i s_i$$

$$E(S) = \frac{3}{12} \times (5) + \dots + \frac{2}{12} \times (15) \quad \underline{E(S) = 9,5 \text{ €}}$$

d) Comparer $E(S)$ avec $E(X) + E(Y)$.

Réponse

$$E(S) = 9,5 \text{ €}$$

D'autre part :

$$E(X) + E(Y) = 7,5 + 2 = 9,5 \text{ €}$$

Donc

$$E(S) = E(X) + E(Y)$$

c'est-à-dire

$E(X + Y) = E(X) + E(Y)$

e) Calculer $V(S)$.

Réponse

$$V(S) = \sum_{i=1}^5 p_i s_i^2 - (E(S))^2$$

$$V(S) = \frac{3}{12} \times (5)^2 + \dots + \frac{2}{12} \times (15)^2 - (9,5)^2 \quad V(S) = 11,25$$

f) Les deux parties sont-elles des expériences aléatoires indépendantes ?

Réponse

Oui, puisque le résultat de la première partie n'intervient pas sur le résultat de la deuxième partie.

g) Comparer $V(S)$ avec $V(X) + V(Y)$.

Réponse

On a précédemment trouvé $V(X) = 6,25$ et $V(Y) = 5$

Donc $V(X) + V(Y) = 6,25 + 5 = 11,25$

Et comme $V(S) = 11,25$, alors

$$V(S) = V(X) + V(Y)$$

c'est-à-dire, que dans le cas de deux **variables aléatoires indépendantes** X et Y on a :

$$V(X + Y) = V(X) + V(Y)$$

2.2 Somme de deux variables aléatoires

X et Y sont deux variables aléatoires.

$S = X + Y$ est la variable qui prend pour valeurs les sommes des valeurs possibles de X et de Y .

On a alors :

$$E(S) = E(X + Y) = E(X) + E(Y) \text{ (linéarité de l'espérance pour l'addition).}$$

Si X et Y sont associées à deux **expériences aléatoires indépendantes** (on dit alors que les variables aléatoires X et Y sont indépendantes : les résultats de l'une n'ont pas d'influence sur les résultats de l'autre) alors $V(S) = V(X + Y) = V(X) + V(Y)$.

Dans le cas où les expériences ne sont pas indépendantes, il se peut que $V(X + Y) \neq V(X) + V(Y)$.

3 Démonstration des formules de l'espérance et de la variance de la loi binomiale

L'objectif de cette partie est de *démontrer* que si la variable X suit une loi binomiale $\mathcal{B}(n, p)$ alors $E(X) = np$ et $V(X) = np(1 - p)$.

Ces formules ont été admises dans le chapitre 4. Loi binomiale.

3.1 Décomposition d'une variable aléatoire X suivant une loi binomiale $\mathcal{B}(n, p)$

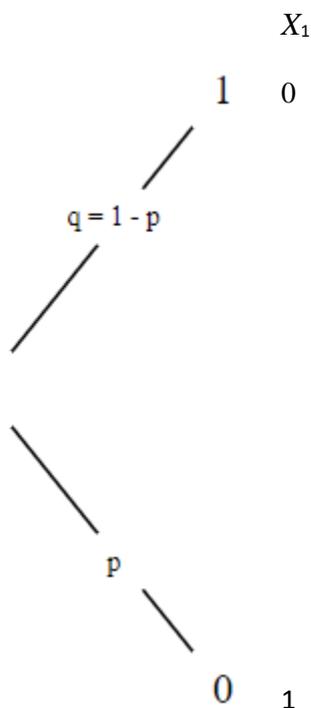
- On décompose la variable aléatoire X comme une somme de variables aléatoires plus simples.
- La variable aléatoire X qui suit la loi binomiale $\mathcal{B}(n, p)$, se décompose en :

$$X = X_1 + X_2 + \dots + X_n$$

où les variables aléatoires X_1, X_2, \dots, X_n sont des **variables aléatoires indépendantes** qui suivent toutes la même loi de Bernoulli $\mathcal{B}(p)$.

3.2 Épreuve de Bernoulli, variable X_1 suivant une loi de Bernoulli

Une *épreuve de Bernoulli* est une épreuve aléatoire à 2 issues « succès » ou « échec ». La probabilité du succès est p .



Si X_1 est une variable aléatoire qui compte le nombre de "succès" lors d'une épreuve de Bernoulli alors sa loi de probabilité est la loi de Bernoulli notée $\mathcal{B}(p)$ donnée par le tableau suivant :

Valeurs de X_1	0	1	TOTAL
$P(X_1 = x_i)$	$q = 1 - p$	p	1

Que valent $E(X_1)$, $V(X_1)$ et $\sigma(X_1)$?

Réponse :

$$E(X_1) = \sum_{i=1}^2 p_i x_i$$

$$E(X_1) = q \times (0) + p \times (1)$$

$$E(X_1) = p$$

Si X_1 suit la loi de Bernoulli de paramètre p , alors :

$$E(X_1) = p$$

$$V(X_1) = \sum_{i=1}^2 p_i x_i^2 - (E(X_1))^2$$

$$V(X_1) = q \times (0)^2 + p \times (1)^2 - (p)^2 \quad V(X_1) = p - p^2$$

Si X_1 suit la loi de Bernoulli de paramètre p , alors :

$$V(X_1) = p(1 - p) = pq$$

$$\sigma(X_1) = \sqrt{V(X_1)} \quad \sigma(X_1) = \sqrt{p(1 - p)} = \sqrt{pq}$$

Si X_1 suit la loi de Bernoulli de paramètre p , alors :

$$\sigma(X_1) = \sqrt{p(1 - p)} = \sqrt{pq}$$

3.3 Formules de l'espérance et de la variance pour une variable aléatoire X qui suit une loi binomiale

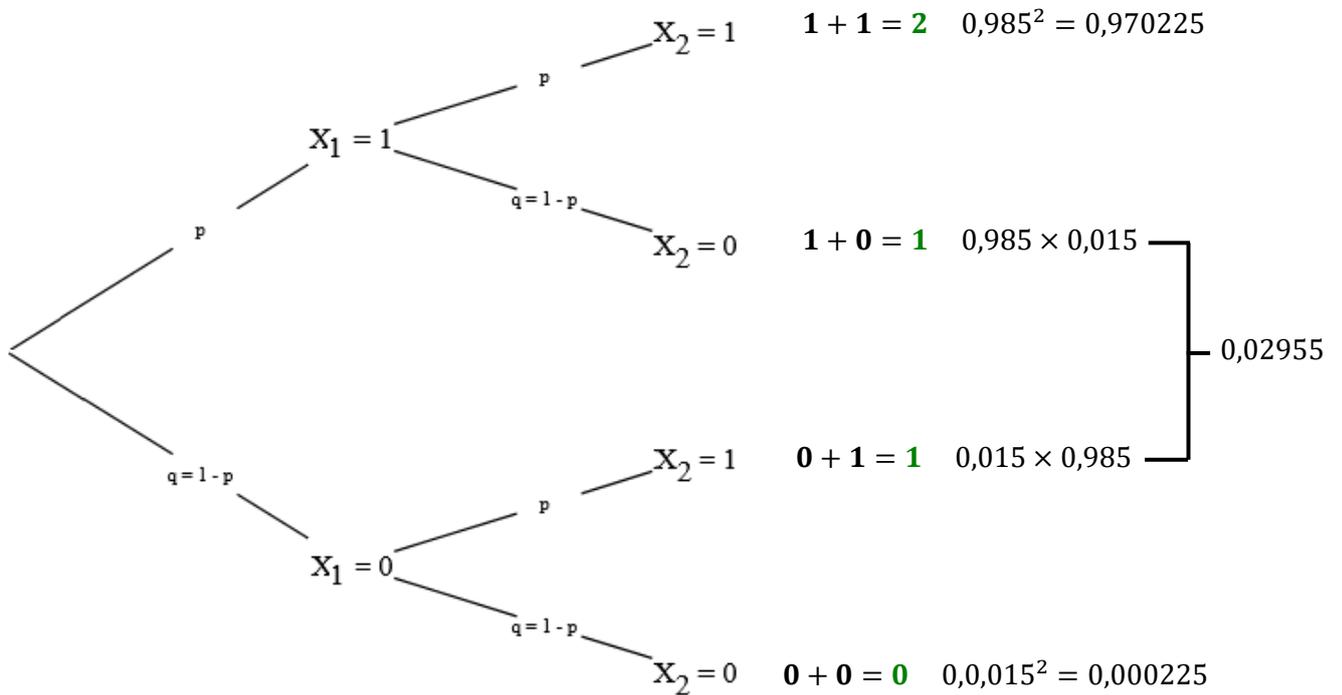
Activité 1

Soit la variable aléatoire S qui suit la loi binomiale $\mathcal{B}(2 ; 0,985)$

On peut considérer deux tirages successifs identiques et indépendants avec comme probabilité du succès $p = 0,985$ et comme probabilité de l'échec $q = 1 - p = 0,015$.

- Aussi on peut considérer que la variable aléatoire S est la somme de deux variables aléatoires X_1 et X_2 qui suivent toutes les deux la loi de Bernoulli $\mathcal{B}(0,985)$. Les variables X_1 et X_2 prennent la valeur 1 en cas de "succès" et 0 en cas d' "échec".

S



On résume dans un tableau toutes les valeurs possibles de la somme S selon les valeurs de X_1 et X_2

Sur la première ligne on a mis toutes les valeurs possibles prises par X_1 .

Sur la première colonne on a mis toutes les valeurs possibles prises par X_2 .

$X_2 \backslash X_1$	0	1
0	0	1
1	1	2

D'où la loi de probabilité de la somme S :

s_i	0	1	2	TOTAL
p_i	0,000225	0,02955	0,970225	1

On retrouve bien la loi binomiale $\mathcal{B}(2 ; 0,985)$

1) Calculer $E(S)$.

Réponse

Au lieu de calculer l'espérance par la formule

$$E(S) = \sum_{i=1}^3 p_i s_i$$

nous utilisons un résultat vu précédemment :

$$E(S) = E(X_1) + E(X_2) \text{ (linéarité de l'espérance).}$$

Les variables aléatoires X_1 et X_2 suivent la loi de Bernoulli de paramètre $p = 0,985$ donc

$$E(X_1) = E(X_2) = p$$

Donc S a pour espérance :

$$E(S) = 2p = 2 \times 0,985$$

$$E(S) = 1,97$$

2) Calculer $V(S)$.

Réponse

Au lieu de calculer l'espérance par la formule

$$V(S) = \sum_{i=1}^3 p_i s_i^2 - (E(S))^2$$

nous utilisons un résultat vu précédemment :

$V(S) = V(X_1) + V(X_2)$ (utilisable lorsque les expériences associées aux variables aléatoires sont indépendantes, ce qui est le cas puisque S suit une loi binomiale).

Les variables aléatoires X_1 et X_2 suivent la loi de Bernoulli de paramètre $p = 0,985$ donc

$$V(X_1) = V(X_2) = pq$$

Donc S a pour espérance :

$$V(S) = 2pq = 2 \times 0,985 \times (1 - 0,985)$$

$$V(S) = 1,94045$$

Activité 2

On considère la variable aléatoire S définie par :

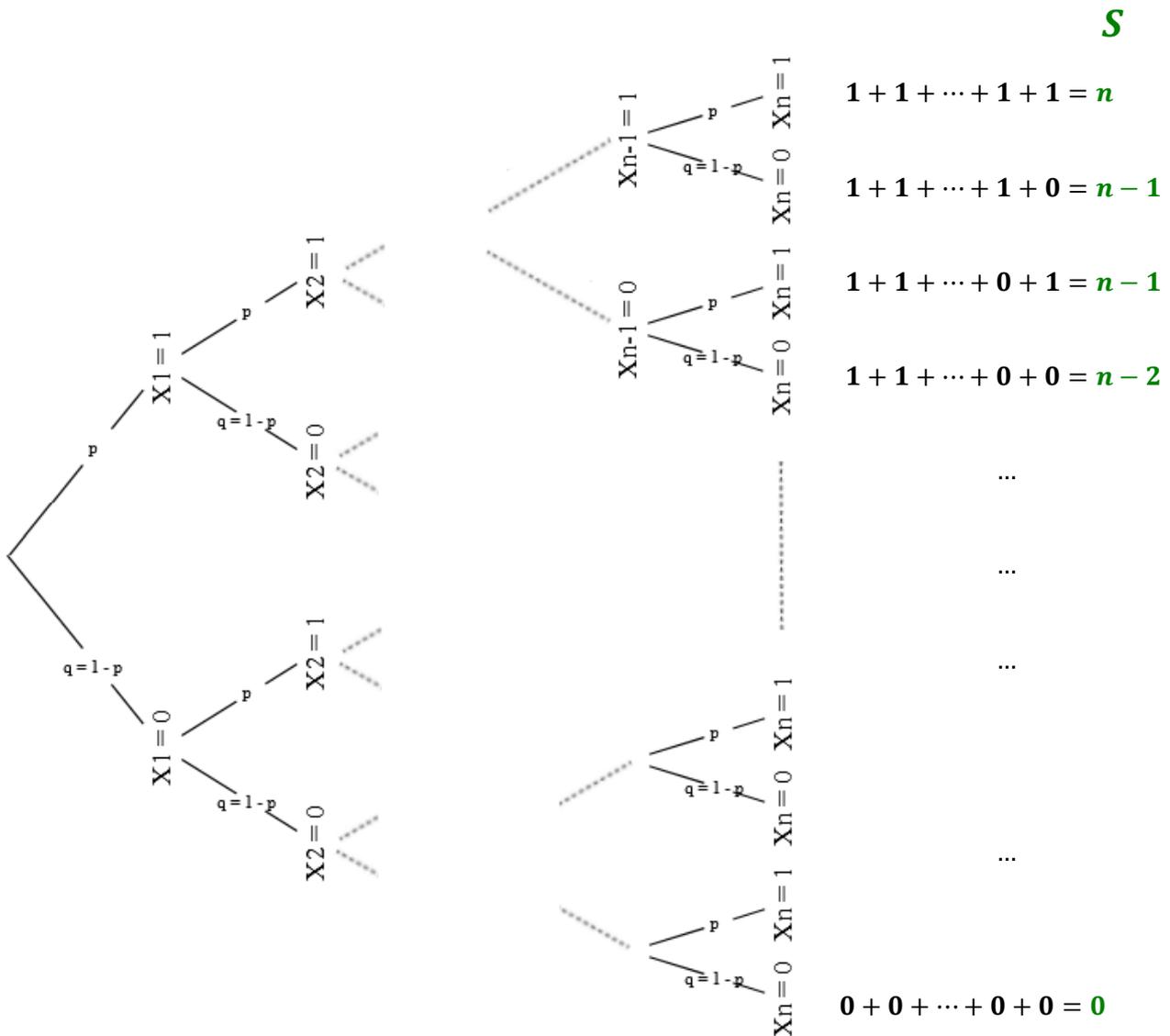
$S = X_1 + X_2 + \dots + X_n$ où les variables aléatoires X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes qui suivent la même loi de Bernoulli $\mathcal{B}(p)$.

X_i est la variable aléatoire qui donne le nombre de succès à la $i^{\text{ème}}$ épreuve. X_i prend la valeur 0 en cas d'échec et la valeur 1 en cas de succès. S donne le nombre de succès au total sur n épreuves.

- 1) Retrouver la formule de $E(S)$ en utilisant la propriété de linéarité de l'espérance.
- 2) Retrouver la formule de $V(S)$ et $\sigma(S)$ en utilisant le fait que S est une somme de variables aléatoires indépendantes.

Réponse :

- 1) Les variables X_1, X_2, \dots, X_n prennent la valeur 1 en cas de "succès" et 0 en cas d' "échec".



$E(S) = E(X_1) + \dots + E(X_n)$ par linéarité de l'espérance.

$$E(S) = \sum_{i=1}^n E(X_i)$$

Les variables aléatoires X_i suivent toutes la loi de Bernoulli de paramètre p (probabilité d'avoir un "succès" à une épreuve) donc :

$$E(X_1) = \dots = E(X_n) = p$$

Donc S a pour espérance :

$$E(S) = \sum_{i=1}^n p$$

$$E(S) = p + p + \dots + p$$

$E(S) = np$

2) S est une somme de **variables aléatoires indépendantes**. Dans ce cas, on a vu que la variance vérifie la propriété :

$$V(S) = V(X_1) + \dots + V(X_n)$$

$$V(S) = \sum_{i=1}^n V(X_i)$$

Les variables aléatoires X_i suivent la loi de Bernoulli de paramètre p (probabilité d'avoir un "succès" à une épreuve) donc :

$$V(X_1) = \dots = V(X_n) = pq$$

Donc S a pour variance :

$$V(S) = \sum_{i=1}^n pq$$

$$V(S) = pq + pq + \dots + pq$$

$V(S) = npq$

Écart-type de la variable aléatoire S qui suit la loi binomiale $\mathcal{B}(n; p)$

$$\sigma(S) = \sqrt{V(S)} \quad \sigma(S) = \sqrt{npq}$$

$\sigma(S) = \sqrt{np(1-p)} = \sqrt{npq}$

4 Échantillon de taille n d'une loi de probabilité X

4.1 Définition d'un échantillon de taille n d'une loi de probabilité sur un exemple

Exemple

- En France en 2018, selon l'INSEE, 75,4 % des individus de 15 à 29 ans ont réalisé un achat sur internet au cours des 12 derniers mois.
- On interroge 500 personnes de la population française, âgées d'entre 15 et 29 ans, pour savoir si elles ont réalisé un achat sur internet au cours des 12 derniers mois.
- Au vu de la taille de la population française, on suppose que les tirages au sort successifs ne changent pas les probabilités que la réponse soit positive ou non, et donc ce prélèvement de 500 personnes par tirage au sort peut être assimilé à un tirage avec remise.
- On considère la liste des variables $(X_1, X_2, \dots, X_{500})$ où X_i vaut 1 si la i -ème personne a réalisé un achat internet au cours des 12 mois et 0 sinon, afin de modéliser l'enquête auprès des 500 personnes.
- **Un échantillon de taille n d'une loi de probabilité est une liste (X_1, X_2, \dots, X_n) de variables aléatoires indépendantes et identiques suivant cette loi.**

Expliquer pourquoi la liste $(X_1, X_2, \dots, X_{500})$ peut être considérée comme un échantillon de variables aléatoires en précisant la loi suivie par les variables aléatoires X_i .

Réponse

Un échantillon de taille $n = 500$ de la population est constitué d'une liste de 500 réponses "1" si la personne a répondu "oui" et "0" si la personne a répondu "non".

Chacune des réponses peut être considérée comme étant la valeur d'une variable aléatoire X_i ne pouvant prendre que deux valeurs 0 ou 1.

Donc l'échantillon de taille $n = 500$ peut être vu comme étant la liste $(X_1, X_2, \dots, X_{500})$ des réponses. Par exemple on obtient la liste $(1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, \dots, 1)$ qui contient 500 nombres traduisant que la 1^{ère} personne a répondu "oui", le 2^e "non", les trois suivantes "oui" etc.

4.2 Espérance, variance et écart-type de la somme $S_n = X_1 + X_2 + \dots + X_n$ et de la moyenne empirique $M_n = S_n/n$ à l'aide de l'exemple précédent

Suite de l'exemple

- 1) Soit $S_{500} = X_1 + X_2 + \dots + X_{500}$.
 - a) Calculer $E(S_{500})$. Exprimer $E(S_{500})$ en fonction de $E(X_1)$.
 - b) Calculer $V(S_{500})$. Exprimer $V(S_{500})$ en fonction de $V(X_1)$.

Réponse

a) Par linéarité de l'espérance, on a :

$$E(S_{500}) = \sum_{i=1}^{500} E(X_i)$$

Or chacune des variables aléatoires X_i suit la même loi de Bernoulli de paramètre $p = 0,754$, donc chacune a la même espérance et cette espérance est égale au paramètre p .

$$E(S_{500}) = \sum_{i=1}^{500} 0,754$$

$$E(S_{500}) = 500 \times 0,754$$

$$E(S_{500}) = 377 \text{ "oui"}$$

b) Puisque les 500 variables aléatoires de la liste $(X_1, X_2, \dots, X_{500})$ sont indépendantes, alors :

$$V(S_{500}) = V(X_1) \dots + V(X_{500})$$

$$V(S_{500}) = \sum_{i=1}^{500} V(X_i)$$

Les variables aléatoires X_i et suivent la loi de Bernoulli de variance pq (p probabilité d'avoir un "succès" à une épreuve et q probabilité d'avoir un "échec") donc :

$$V(X_1) = \dots = V(X_{500}) = pq$$

Donc S_{500} a pour variance :

$$V(S_{500}) = \sum_{i=1}^{500} pq$$

$$V(S_{500}) = 500(pq)$$

$$V(S_{500}) = 500 \times 0,754 \times (1 - 0,754)$$

$$V(S_{500}) = 92,742$$

2) On considère la **moyenne "empirique" des 500 variables aléatoires X_i**

$$M_{500} = \frac{S_{500}}{500}$$

Elle représente la probabilité de réponse "oui" par individu.

a) Calculer $E(M_{500})$. Exprimer $E(M_{500})$ en fonction de $E(X_1)$.

b) Calculer $V(M_{500})$. Exprimer $V(M_{500})$ en fonction de $V(X_1)$.

Réponse

Puisque $M_{500} = \frac{S_{500}}{500}$ alors on peut donc écrire :

$$M_{500} = \frac{1}{500} \times S_{500}$$

Nous allons considérer que M_{500} est une transformation affine de la variable aléatoire S_{500} .

Nous pouvons donc utiliser les propriétés :

Si $M_{500} = aS_{500} + b$ alors on a :

$$E(M_{500}) = E(aS_{500} + b) = aE(S_{500}) + b$$

$$V(M_{500}) = V(aS_{500} + b) = V(aS_{500}) = a^2V(S_{500}).$$

a)

$$E(M_{500}) = E\left(\frac{1}{500}S_{500} + 0\right) = \frac{1}{500}E(S_{500})$$

On a déjà calculé que

$$E(S_{500}) = 500 \times E(X_1)$$

donc :

$$E(M_{500}) = \frac{1}{500}500 \times E(X_1)$$

$$E(M_{500}) = E(X_1)$$

b)

$$V(M_{500}) = V\left(\frac{1}{500}S_{500} + 0\right) = \left(\frac{1}{500}\right)^2 V(S_{500})$$

On a déjà calculé que

$$V(S_{500}) = 500 \times V(X_1)$$

donc :

$$V(M_{500}) = \frac{1}{500^2}500 \times V(X_1)$$

$$V(M_{500}) = \frac{1}{500}V(X_1)$$

4.3 Définition d'un échantillon d'une loi de probabilité

On appelle **échantillon** de taille n d'une loi de probabilité toute **liste** (X_1, X_2, \dots, X_n) de variables aléatoires indépendantes qui suivent toute cette loi de probabilité.

La somme de l'échantillon est une autre variable aléatoire $S_n = X_1 + X_2 + \dots + X_n$.

La moyenne de l'échantillon est encore une autre variable aléatoire

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Remarque

Dans un échantillon d'une loi de probabilité, toutes les variables aléatoires X_i de la liste ont les mêmes espérances, variance et écart-type.

Par exemple, si on prend X_1 la première des variables aléatoires de l'échantillon de la loi de probabilité, on a :

$$E(S_n) = n \times E(X_1)$$

$$V(S_n) = n \times V(X_1)$$

$$\sigma(S_n) = \sqrt{n} \times \sigma(X_1)$$

$$E(M_n) = E(X_1)$$

$$V(M_n) = \frac{V(X_1)}{n}$$

$$\sigma(M_n) = \frac{\sigma(X_1)}{\sqrt{n}}$$

Exemple : Une roue de loterie comporte cinq secteurs angulaires égaux. Les deux premiers secteurs valent 300 points, le troisième vaut 100 points et les deux derniers valent -400 points.

On fait tourner la roue quatre fois de suite et on gagne la somme S_4 des points cumulés sur les quatre lancers de la roue.

- 1) Décomposer S_4 en une somme de variables aléatoires identiques et indépendantes dont on donnera la loi.
- 2) Calculer $E(S_4)$, $V(S_4)$ au centième près.
- 3) On pose $M_4 = \frac{S_4}{4}$ la moyenne empirique¹ de l'échantillon. Déterminer au centième près l'espérance et la variance de M_4 .

¹ **Empirique** : Qui se base uniquement sur les observations.

Réponse

1)

$$S_4 = X_1 + X_2 + X_3 + X_4$$

où chaque variable aléatoire X_i donne le nombre de points au i -ème lancer.

Les X_i sont indépendantes car associées à des épreuves indépendantes.

La loi suivie par ces variables X_i est :

x_i	-400	100	300	TOTAL
p_i	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

2) On commence par calculer l'espérance et la variance d'une des variables aléatoires, par exemple X_1 .

$$E(X_1) = \sum_{i=1}^3 p_i x_i$$

$$E(X_1) = \frac{2}{5} \times (-400) + \frac{1}{5} \times (100) + \frac{2}{5} \times (300) \qquad E(X_1) = -20 \text{ points}$$

et comme

$$E(S_4) = 4 \times E(X_1)$$

alors

$$\underline{E(S_4) = 4 \times -20 = -80 \text{ points}}$$

$$V(X_1) = \sum_{i=1}^3 p_i x_i^2 - (E(X_1))^2$$

$$V(X_1) = \frac{2}{5} \times (-400)^2 + \frac{1}{5} \times (100)^2 + \frac{2}{5} \times (300)^2 - (-20)^2 \qquad V(X_1) = 101600$$

et comme

$$V(S_4) = 4 \times V(X_1)$$

alors

$$\underline{V(S_4) = 4 \times 101600 = 406400}$$

3) La moyenne empirique de l'échantillon est

$$M_4 = \frac{S_4}{4}$$

$$E(M_4) = E(X_1) = -20 \text{ points}$$

L'espérance de la moyenne empirique est -20 points.

$$V(M_4) = \frac{1}{4} V(X_1)$$

$$V(M_4) = \frac{1}{4} \times 101600 = 25400$$

La variance de la moyenne empirique est 25400.

5 Inégalité de Bienaymé-Tchebychev

5.1 Inégalité de Bienaymé-Tchebychev sur un exemple

L'inégalité de Bienaymé-Tchebychev est un outil pour étudier la dispersion de la variable aléatoire X autour de son espérance mathématique $E(X)$.

Irénée-Jules Bienaymé (1796-1878) est un mathématicien français qui a énoncé cette inégalité en 1853. **Pafnouti Tchebychev** (1821-1894) est un mathématicien russe qui l'a démontrée en 1867.

Exemple

On considère la variable aléatoire X donnant le gain en euros à un jeu de grattage dont la loi de probabilité est donnée ci-dessous :

x_i	0	2	4	6	10	20	100	1000	25000
p_i	0,682	0,1435	0,103	0,0363	0,0225	0,0126	$9,73 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	$7 \cdot 10^{-7}$

1) Déterminer $E(X)$.

Pour la suite on prendra la valeur de $E(X)$ arrondie à 10^{-3} .

Réponse

$$E(X) = \sum_{i=1}^9 p_i x_i$$

$$E(X) = 0,682 \times (0) + \dots + 7 \cdot 10^{-7} \times (25000)$$

$$E(X) \approx 1,423 \text{ €}$$

2) Déterminer les probabilités suivantes :

a) $P(|X - E(X)| \geq 15)$

b) $P(|X - E(X)| \geq 10000)$

Réponse

La valeur absolue de la différence de deux nombres a et b est notée $|a - b|$.

Elle est toujours positive et représente la distance entre les deux nombres.

Exemples :

$|12 - 10| = |2| = 2$ est la distance entre 10 et 12.

$|10 - 12| = |-2| = 2$ est aussi la distance entre 10 et 12.

- a) Donc la première probabilité à calculer $P(|X - E(X)| \geq 15)$ est la probabilité pour que la valeur de X ait une distance par rapport à $E(X) = 1,423$ supérieure ou égale à 15

Or les valeurs x_i possibles sont 0, 2, 4, 6, 10, 20, 100, 1000, 25000.

Si on résume les distances $|x_i - 1,423|$ dans un tableau on obtient :

x_i	0	2	4	6	10	20	100	1000	25000
$ x_i - 1,423 $	1,423	0,577	2,577	4,577	8,577	18,577	98,577	998,58	24999
p_i	0,682	0,1435	0,103	0,0363	0,0225	0,0126	$9,73 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	$7 \cdot 10^{-7}$

Seulement quatre valeurs x_i vérifient la condition "distance supérieure ou égale à 15", totalisant les probabilités $0,0126$; $9,73 \cdot 10^{-5}$; $2 \cdot 10^{-6}$; $7 \cdot 10^{-7}$ soit un total de 0,0127.

Conclusion : $P(|X - E(X)| \geq 15) = 0,0127$

La probabilité que le gain au grattage s'écarte de plus de 15 € par rapport à l'espérance 1,423 € est de 0,0127 soit dans 1,27% des cas.

- b) La deuxième probabilité à calculer $P(|X - E(X)| \geq 10000)$ est la probabilité pour que la valeur de X ait une distance par rapport à $E(X) = 1,423$ supérieure ou égale à 10000

Or les valeurs x_i possibles sont 0, 2, 4, 6, 10, 20, 100, 1000, 25000.

Si on résume les distances $|x_i - 1,423|$ dans un tableau on obtient :

x_i	0	2	4	6	10	20	100	1000	25000
$ x_i - 1,423 $	1,423	0,577	2,577	4,577	8,577	18,577	98,577	998,58	24999
p_i	0,682	0,1435	0,103	0,0363	0,0225	0,0126	$9,73 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	$7 \cdot 10^{-7}$

Seule une valeur x_i vérifie la condition "distance supérieure ou égale à 10000", correspondant à la probabilité 7.10^{-7} .

Conclusion : $P(|X - E(X)| \geq 10000) = 7 \times 10^{-7}$

La probabilité que le gain au grattage s'écarte de plus de 10000 € par rapport à l'espérance 1,423 € est de 7.10^{-7} soit dans 0,7 millionième des cas.

3) Soit t un réel *strictement positif*.

Que semble-t-on pouvoir dire de la probabilité $P(|X - E(X)| \geq t)$ quand t devient grand ?

Réponse

Si on résume les deux résultats de la question précédente :

Valeur de t	Valeur de $p(X - E(X) \geq t)$
15	0,0127
10000	7.10^{-7}

Il semble que **lorsque t devient grand**, la **probabilité** pour que x_i s'écarte de plus de t par rapport à l'espérance **tende vers 0**.

4) Calculer $V(X)$. Pour la suite on prendra la valeur de $V(X)$ arrondie à l'entier.

Réponse

$$V(X) = \sum_{i=1}^9 p_i x_i^2 - (E(x))^2$$

$$V(X) = 0,682 \times (0)^2 + \dots + 7.10^{-7} \times (25000)^2 - 1,423^2 \quad \underline{V(X) \approx 449}$$

5) Calculer $\frac{V(X)}{t^2}$ pour $t = 15$ et pour $t = 10000$.

Réponse

$$V(X) = 449$$

- Pour $t = 15$:

$$\frac{V(X)}{t^2} = \frac{449}{15^2}$$

$$\frac{V(X)}{t^2} \approx 1,996$$

- Pour $t = 10000$:

$$\frac{V(X)}{t^2} = \frac{449}{10000^2}$$

$$\frac{V(X)}{t^2} \approx 4,49 \cdot 10^{-6}$$

6) Comparer $P(|X - E(X)| \geq t)$ et $\frac{V(X)}{t^2}$ dans ces deux cas.

Réponse

	t	$P(X - E(X) \geq t)$	$\frac{V(X)}{t^2}$	Comparaison
1 ^{er} cas	15	0,0127	1,996	$P(X - E(X) \geq t)$ est inférieure à $\frac{V(X)}{t^2}$
2 ^e cas	10000	7×10^{-7}	$4,49 \cdot 10^{-6}$	$P(X - E(X) \geq t)$ est inférieure à $\frac{V(X)}{t^2}$

7) Quelle conjecture peut-on faire sur la comparaison de $P(|X - E(X)| \geq t)$ avec $\frac{V(X)}{t^2}$?

Réponse

En se basant sur les deux cas ($t = 15$ et $t = 10000$) on conjecture que, quel que soit le réel $t > 0$,

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}$$

5.2 Formule de l'inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

Pour tout réel strictement positif t , on a :

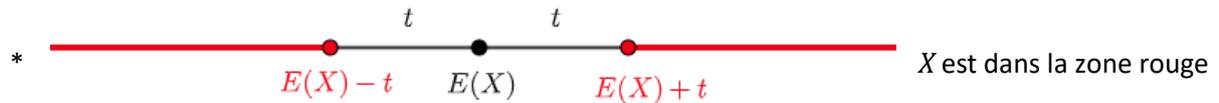
$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}$$

Ceci est l'inégalité de Bienaymé-Tchebychev.

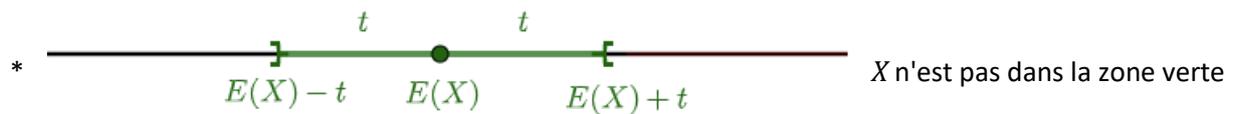
Quand on connaît la variance d'une variable aléatoire X , alors on a un *majorant* de la probabilité $P(|X - E(X)| \geq t)$. Ce majorant est $\frac{V(X)}{t^2}$.

Remarques

- * $|X - E(X)| \geq t$ **équivalent** successivement à :
- * La distance entre X et son espérance est supérieure ou égale à t



* $X \in]-\infty ; E(X) - t] \cup [E(X) + t ; +\infty [$



* $X \notin]E(X) - t ; E(X) + t [$

- Il ressort de cela que $X \in]E(X) - t ; E(X) + t [$ est l'**évènement contraire** de

$$X \in]-\infty ; E(X) - t] \cup [E(X) + t ; +\infty [$$

et donc en utilisant la probabilité de l'évènement contraire :

$$P(X \in]E(X) - t ; E(X) + t [) = \mathbf{1} - P(X \in]-\infty ; E(X) - t] \cup [E(X) + t ; +\infty [)$$

Conséquence de l'inégalité de Bienaymé-Tchebychev :

L'inégalité de Bienaymé-Tchebychev s'écrit :

$$\forall t \in \mathbb{R}^{+*}, P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}$$

$$-P(|X - E(X)| \geq t) \geq -\frac{V(X)}{t^2}$$

$$1 - P(|X - E(X)| \geq t) \geq 1 - \frac{V(X)}{t^2}$$

$$P(|X - E(X)| < t) \geq 1 - \frac{V(X)}{t^2}$$

$$P(X \in]E(X) - t ; E(X) + t [) \geq 1 - \frac{V(X)}{t^2}$$

On dit que $]E(X) - t ; E(X) + t [$ est un **intervalle de fluctuation de X centré sur son espérance**.

Cette inégalité permet de *minorer* la probabilité que X appartienne à cet intervalle de fluctuation.

5.3 Exemple d'utilisation de l'inégalité de Bienaymé-Tchebychev

Dans une usine, la variable aléatoire L donnant la largeur en millimètres d'une puce électronique prise au hasard a pour espérance $E(L) = 12$ et pour variance $V(L) = 0,01$.

Si la largeur d'une puce n'appartient pas à $]11; 13[$, alors la puce n'est pas commercialisable. A l'aide de l'inégalité de Bienaymé-Tchebychev répondre aux questions suivantes :

- a) Montrer que la probabilité que la puce n'est pas commercialisable est inférieure ou égale à 0,01.

Réponse : La puce n'est pas commercialisable lorsque $L \notin]11; 13[$.

$L \notin]11; 13[$ équivaut à $L \in]-\infty; 11] \cup [13; +\infty[$

La puce n'est pas commercialisable lorsque la distance entre L et 12 est supérieure ou égale à 1.

L'inégalité de Bienaymé-Tchebychev donne une majoration de la probabilité de $P(|L - 12| \geq 1)$.

$$P(|L - 12| \geq 1) \leq \frac{V(L)}{t^2}$$

Ici $t = 1$ et $V(L) = 0,01$, donc :

$$P(|L - 12| \geq 1) \leq \frac{0,01}{1^2}$$

$$P(|L - 12| \geq 1) \leq 0,01$$

Remarque: Cela peut aussi s'écrire $P(L \in]-\infty; 11] \cup [13; +\infty[) \leq 0,01$

Conclusion : La probabilité que la puce ne soit pas commercialisable est inférieure ou égale à 0,01.

- b) Majorer la probabilité que la largeur de la puce soit supérieure ou égale à 13 mm.

Réponse : L'inégalité de Bienaymé-Tchebychev a permis dans la question a) de montrer que la probabilité que la valeur de L se trouve dans $] -\infty ; 11] \cup [13 ; +\infty[$ est inférieure ou égale à 0,01.

Donc la probabilité que la valeur de L se trouve dans $[13 ; +\infty[$ est aussi inférieure ou égale à 0,01.

D'où la majoration : $P(L \geq 13) \leq 0,01$.

5.4 Cas de la loi binomiale

Exemple : X suit la loi binomiale $\mathcal{B}(20; 0,45)$.

- 1) Donner la majoration donnée par l'inégalité de Bienaymé-Tchebychev de la probabilité :

$$P(|X - E(X)| \geq 2 \sigma(X)).$$

- 2) Calculer cette même probabilité avec la calculatrice.
3) Que penser de la majoration donnée par l'inégalité de Bienaymé-Tchebychev ?

Réponse

1) Énoncé de l'inégalité de Bienaymé-Tchebychev :

$$\forall t \in \mathbb{R}^{+*}, \quad P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}$$

Dans cet exercice on a $t = 2\sigma(X)$, donc l'inégalité s'écrit :

$$P(|X - E(X)| \geq 2\sigma(X)) \leq \frac{V(X)}{(2\sigma(X))^2}$$

$$P(|X - E(X)| \geq 2\sigma(X)) \leq \frac{V(X)}{4\sigma(X)^2}$$

$$P(|X - E(X)| \geq 2\sigma(X)) \leq \frac{V(X)}{4V(X)}$$

$$P(|X - E(X)| \geq 2\sigma(X)) \leq \frac{1}{4}$$

- Application numérique au cas où X suit la loi binomiale $\mathcal{B}(20 ; 0,45)$:

$$n = 20, p = 0,45, q = 0,55$$

On sait que si X suit une loi binomiale, alors $E(X) = np$ et $V(X) = npq$.

$$\text{Donc } E(X) = 20 \times 0,45 = 9$$

$$\text{et } V(X) = 20 \times 0,45 \times 0,55 = 4,95; \text{ déduit que } \sigma(X) = \sqrt{4,95} \approx 2,225 \text{ donc } \mathbf{2\sigma = 4,45}$$

Donc l'inégalité de Bienaymé-Tchebychev permet d'affirmer que $p(|X - 9| \geq 2 \times 2,225) \leq \frac{1}{4}$

$$P(|X - 9| \geq 4,45) \leq \frac{1}{4}$$

2) On fait le calcul exact à la calculatrice. Pour mieux se rendre compte, on peut tracer sur la calculatrice la représentation graphique des probabilités de la loi binomiale $\mathcal{B}(20 ; 0,45)$.

Voici la procédure détaillée sur la TI-83 :

4:EffListe pour avoir L₁ la touche (au-dessus du "7"), pour avoir L₂

1:Modifier

- Remplissage de la liste L₁ avec les valeurs x_i

Sélectionner le **titre** de la liste L₁ et appuyer sur

Sous le tableau, le curseur clignote à côté de L₁=

La loi binomiale ici a pour paramètre $n = 20$, donc les valeurs de la variable aléatoire discrète x_i peuvent aller de 0 à 20. Pour entrer rapidement le liste des entiers de 0 à 20, procéder ainsi :

choisir le menu OP

Sélectionner 5:suite

Expr:X

Variable:X

début:0

fin:20

pas:1

Sélectionner Coller et appuyer sur

Appuyer encore sur . La liste L_1 se remplit des entiers de 0 à 20

- Remplissage de la liste L_2 avec les valeurs p_i

Sélectionner **le titre** de la liste L_2 et appuyer sur

Sous le tableau, le curseur clignote à côté de $L_2=$

choisir la distribution **binomFdp**(et appuyer sur

nbreEssais:20

p:0.45

valeur de x: (pour sélectionner la liste L_1)

Sélectionner Coller et appuyer sur

Appuyer encore sur . La liste L_2 se remplit des probabilités p_i que X prenne telle ou telle valeur entière x_i entre 0 et 20 incluses, selon la loi binomiale $\mathcal{B}(20 ; 0.45)$

- Visualisation de la distribution

Appuyer sur la touche et désactiver les représentations de fonctions Y_1, Y_2 etc. qui seraient éventuellement actives.

(Pour les désactiver, il faut sélectionner les signes " = " qui sont sur fond noir et appuyer sur la touche . Le fond noir disparaît ce qui signifie que la représentation graphique de cette fonction est désactivée).

Appuyer sur 2^{nd} graph stats

Se placer sur 1:Graph1 ... NAff et appuyer sur entrer

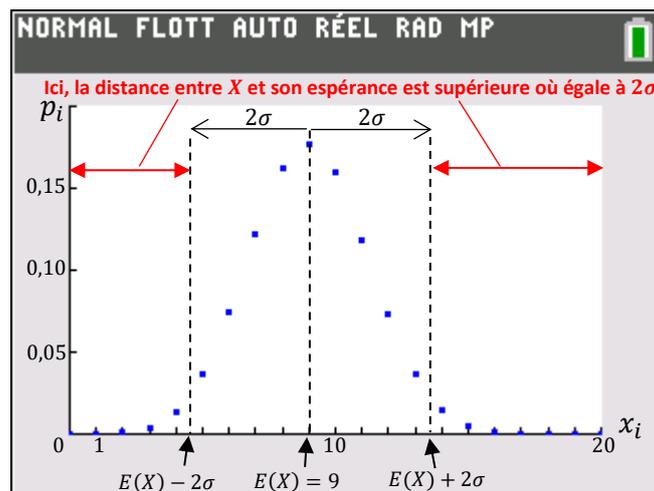
- Sélectionner **Aff** et appuyer sur entrer pour permettre d'afficher le graphique statistique n°1
- Type : sélectionner la première option (des points)
- Xliste :L₁ (si elle n'est pas déjà présente il faut faire 2^{nd} 1 entrer)
- Yliste :L₂ (si elle n'est pas déjà présente il faut faire 2^{nd} 2 entrer)
- Marque : la troisième sorte (des gros points) entrer
- Couleur: à vous de choisir avec les flèches de direction gauche et droite.

Appuyer sur la touche fenêtre

Puisque les x_i vont de 0 à 20, choisissez Xmin=0, Xmax=20, Xgrad=1

Puisque les p_i sont entre 0 et 1, mais que probablement qu'aucun ne sera supérieur à 0.2, choisissez Ymin=0, Ymax=0.2, Ygrad=0.05

Appuyez sur la touche graphe



L'inégalité de Bienaymé-Tchebychev permet d'affirmer que la probabilité que X prenne l'une des valeurs dans la zone rouge sur l'écran de la calculatrice (c'est-à-dire $X \leq 4$ ou $X \geq 14$) est majorée par $\frac{1}{4}$. On va calculer $P(X \leq 4) + P(X \geq 14)$.

Calcul exact avec la loi binomiale sur la calculatrice :

- Pour calculer $P(X \leq 4)$, on utilise :

2^{nd} distrib choisir la distribution **binomFRép**(et appuyer sur entrer

nbreEssais:20

p:0.45

valeur de x:4

Sélectionner Coller et appuyer sur entrer

On voit sur l'écran binomFRép(20,0.45,4)

appuyer sur . On trouve **0,01886** environ.

On stocke ce résultat dans la mémoire A. Appuyer sur A

Le cumul des probabilités $P(X = 0) + \dots + P(X = 4)$ est de l'ordre de 0,019.

- Pour calculer $P(X \geq 14)$, on utilise $P(X \geq 14) = 1 - P(X \leq 13)$.

1 - binomFRép(20,0.45,13) donne environ **0,02141**

On stocke ce résultat dans la mémoire B. Appuyer sur B

Enfin, on additionne pour trouver $P(X \leq 4) + P(X \geq 14)$.

A + B donne environ **0,040**

Donc on peut dire qu'en faisant le calcul avec la loi binomiale, on obtient une probabilité $p(|X - E(X)| \geq 2 \sigma(X))$ **inférieure à 0,041**

- 3) Ce qu'on peut penser de l'inégalité de Bienaymé-Tchebychev : Elle garantissait que $P(|X - E(X)| \geq 2 \sigma(X))$ est **inférieure à 0,25**. Le calcul exact montre que c'est bien le cas puisque $0,041 < 0,25$. **Cependant on voit que la majoration par 0,25 est très imprécise.**

Remarque : Le caractère universel de l'inégalité de Bienaymé-Tchebychev a pour contrepartie le fait qu'elle est loin d'être optimale.

6 Inégalité de concentration de la moyenne empirique

On considère une expérience aléatoire et X est la variable aléatoire associée à cette expérience, d'espérance $E(X)$ et de variance $V(X)$. **On répète n fois cette expérience de manière indépendante.** On obtient un échantillon de n variables aléatoires X_1, X_2, \dots, X_n .

Les variables X_1, X_2, \dots, X_n suivent la même loi de probabilité. On note $E(X)$ leur espérance et $V(X)$ leur variance.

$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ est la variable aléatoire "moyenne empirique" de cet échantillon de taille n .

Pour tout réel strictement positif t , on a :

$$P(|M_n - E(X)| \geq t) \leq \frac{V(X)}{n \times t^2}.$$

L'inégalité de concentration est un outil pour étudier la dispersion de la moyenne empirique d'un échantillon M_n autour de l'espérance mathématique $E(X)$.

Exemple 1

100 personnes jouent indépendamment à un même jeu dont la variable associée au gain (en euros) a pour espérance 10 et pour variance 2.

- 1) Donner une majoration de la probabilité que la *moyenne* des gains de ces 100 personnes soit à une distance supérieure ou égale à 3 par rapport à $E(X)$.
- 2) Donner une minoration de la probabilité que la *moyenne* des gains de ces 100 personnes soit comprise strictement entre 7 € et 13 €.

Réponse

- 1) Notons M_{100} le gain moyen par personne dans le groupe de 100 joueurs. Les $n = 100$ personnes jouent de **manière indépendante**. Donc on peut utiliser l'inégalité de concentration de la moyenne qui donne la majoration de probabilité suivante :

$$P(|M_{100} - E(X)| \geq t) \leq \frac{V(X)}{100 \times t^2}$$

avec $E(X) = 10$, $t = 3$, $V(X) = 2$

$$P(|M_{100} - 10| \geq 3) \leq \frac{2}{100 \times 3^2}$$

$$P(|M_{100} - 10| \geq 3) \leq \frac{2}{900}$$

$$P(|M_{100} - 10| \geq 3) \leq 2,2 \cdot 10^{-3}$$

Conclusion :

La probabilité que la *moyenne empirique* M_{100} des gains de ces 100 personnes soit à une distance supérieure ou égale à 3 par rapport à $E(X) = 10$, c'est-à-dire dans les intervalles $[0 ; 7] \cup [13 ; +\infty[$ est majorée par $2,2 \times 10^{-3}$.

- 2) "La *moyenne* M_{100} des gains de ces 100 personnes est comprise strictement entre 7 € et 13 € " est l'évènement contraire de " la *moyenne* M_{100} est dans les intervalles $[0 ; 7] \cup [13 ; +\infty[$ ".

Puisque la probabilité que M_{100} se trouve dans $[0 ; 7] \cup [13 ; +\infty[$ est plus petite que $\frac{2}{(300)^2}$, alors on en déduit que la probabilité que M_{100} se trouve dans $]7 ; 13[$ est plus grande que

$$1 - \frac{2}{900} \approx 0,9978$$

$$P(|M_{100} - 10| < 3) > 0,9978$$

ce qui équivaut à $P(M_{100} \in]7 ; 13[) > 0,9978$

Exemple 2 Utiliser l'inégalité de concentration pour trouver la taille d'un échantillon

Une urne contient deux jetons sur lesquels figure le nombre 3, deux jetons sur lesquels figure le nombre 5 et un jeton sur lequel figure le nombre 10.

On tire un jeton dans cette urne et on considère la variable aléatoire X donnant le nombre obtenu.

- 1) Déterminer $E(X)$ et $V(X)$.
- 2) Calculer le nombre n de tirages avec remise qu'il faut effectuer dans cette urne pour être sûr au risque de 5 % (ou au seuil de 95 %) que la moyenne M_n des nombres obtenus soit comprise entre 5 et 5,4 exclus.

Réponse

- 1) On commence par écrire la loi de probabilité de la variable aléatoire X qui donne la valeur du nombre tiré de l'urne.

x_i	3	5	10	TOTAL
p_i	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	1

Calculer l'espérance $E(X)$ la variance $V(X)$

$$E(X) = \sum_{i=1}^3 p_i x_i$$

$$E(X) = \frac{2}{5} \times (3) + \frac{2}{5} \times (5) + \frac{1}{5} \times (10) \quad \underline{E(X) = 5,2}$$

$$V(X) = \sum_{i=1}^3 p_i x_i^2 - (E(X))^2$$

$$V(X) = \frac{2}{5} \times (3)^2 + \frac{2}{5} \times (5)^2 + \frac{1}{5} \times (10)^2 - (5,2)^2 \quad \underline{V(X) = 6,56}$$

- 2) Les n tirages se font de **manière indépendante puisqu'ils sont avec remise**. Donc on peut utiliser l'inégalité de concentration de la moyenne qui donne la majoration suivante :

$$\forall t \in \mathbb{R}^{+*}, \quad P(|M_n - E(X)| \geq t) \leq \frac{V(X)}{n \times t^2}$$

avec $E(X) = 5,2$ $t = 0,2$ $V(X) = 6,56$

$$P(|M_n - 5,2| \geq 0,2) \leq \frac{6,56}{n \times 0,2^2}$$

$$P(|M_n - 5,2| \geq 0,2) \leq \frac{164}{n}.$$

La probabilité que la *moyenne* M_n des nombres de ces n tirages soit à une distance supérieure ou égale à 0,2 par rapport à $E(X) = 5,2$, c'est-à-dire dans les intervalles $[0 ; 5] \cup [5,4 ; +\infty[$ est majorée par $\frac{164}{n}$.

- "La *moyenne* M_n des nombres de ces n tirages est comprise strictement entre 5 et 5,4 " est l'évènement contraire de " la *moyenne* M_n est dans les intervalles $[0 ; 5] \cup [5,4 ; +\infty[$ ".

Puisque la probabilité que M_n se trouve dans $[0 ; 5] \cup [5,4 ; +\infty[$ est plus petite que $\frac{164}{n}$, alors on en déduit que la probabilité que M_n se trouve dans $]5 ; 5,4[$ est plus grande que

$$1 - \frac{164}{n}$$

$$P(|M_n - 5,2| < 0,2) \geq 1 - \frac{164}{n}$$

Ceci peut aussi s'écrire :

$$P(M_n \in]5 ; 5,4[) \geq 1 - \frac{164}{n}$$

La probabilité que la moyenne des nombres tirés soit strictement entre 5 et 5,4 est supérieure ou égale à $1 - \frac{164}{n}$.

On cherche n pour que la probabilité soit égale à 0,95.

Donc on résout l'équation : $1 - \frac{164}{n} = 0,95$.

$$-\frac{164}{n} = 0,95 - 1$$

$$\frac{164}{n} = 0,05$$

$$\frac{164}{0,05} = n$$

$$n = 3280$$

Conclusion :

En effectuant 3280 de tirages avec remise, on est sûr au risque de 5 % (ou au seuil de 95 %) que la moyenne M_{3280} des nombres obtenus soit comprise entre 5 et 5,4 exclus.

7 Loi des grands nombres

Soit (X_1, X_2, \dots, X_n) un échantillon de variables aléatoires d'espérance $E(X)$ et $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ la variable aléatoire moyenne de cet échantillon.

D'après l'inégalité de concentration, on a :

$$\forall t \in \mathbb{R}^+, \quad P(|M_n - E(X)| \geq t) \leq \frac{V(X)}{n \times t^2}.$$

Or,

$$\lim_{n \rightarrow +\infty} \left(\frac{V(X)}{n \times t^2} \right) = \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \times \frac{V(X)}{t^2} \right) = 0$$

donc, comme une probabilité est toujours positive :

$$\forall t \in \mathbb{R}^+, \quad \lim_{n \rightarrow +\infty} (P(|M_n - E(X)| \geq t)) = 0$$

En utilisant l'évènement contraire, on peut écrire de façon équivalente :

$$\forall t \in \mathbb{R}^+, \quad \lim_{n \rightarrow +\infty} (P(M_n \in]E(X) - t ; E(X) + t[)) = 1$$

La loi des grands nombres peut s'énoncer ainsi :

La moyenne empirique M_n , calculée sur les valeurs d'un échantillon de n variables aléatoires identiques X_i , converge vers l'espérance $E(X)$ lorsque la taille n de l'échantillon tend vers $+\infty$.

M_n , pour n assez grand, permet d'estimer l'espérance de X .

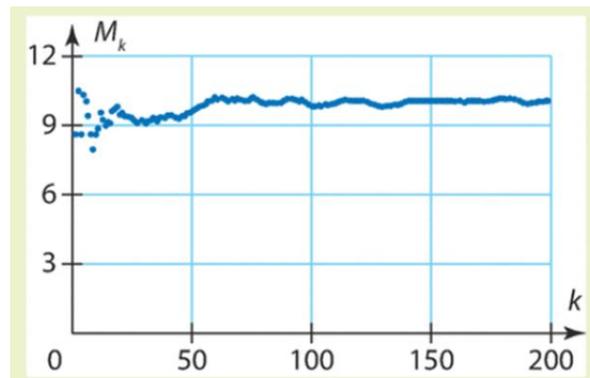
Exemple

On considère un échantillon de variables aléatoires $(X_1, X_2, \dots, X_{200})$ d'espérance $E(X)$.

On considère

$$M_k = \frac{X_1 + X_2 + \dots + X_k}{k}$$

la variable aléatoire moyenne de l'échantillon (X_1, X_2, \dots, X_k) pour k entier entre 1 et 200. On donne ci-contre le nuage de points $(k ; M_k)$.



Estimer $E(X)$.

Réponse

On observe que les ordonnées M_k des points tendent vers 10 lorsque k devient grand.

Donc, d'après la loi des grands nombres, on estime que $E(X) = 10$.