Somme de variables aléatoires, concentration, loi des grands nombres

Table des matières

-

1 Somme de deux variables aléatoires

1.1 Définition

Définition 1 : Soit X et Y deux variables aléatoires associées à une même expérience d'univers fini Ω et a un réel.

$X+Y$ et aX sont deux variables aléatoires définies sur Ω qui prennent comme valeur pour un événement donné respectivement : la somme des valeurs de X et Y et le produit de a par X.

Exemple : On lance deux dés, l'un tétraédrique numéroté de 1 à 4 et l'autre cubique numéroté de 1 à 6. On appelle X et Y les variables aléatoires associées respectivement aux résultats du dé tétraédrique et du dé cubique.

- $X+Y$ est la variable aléatoire qui prend les valeurs de 2 à 10.
- $2X$ est la variable aléatoire qui prend les valeurs 2, 4, 6, 8.

Remarque : On peut généraliser la somme de variable aléatoires à n variables. Par exemple : on lance 3 dés cubiques de couleurs différentes et l'on note X, Y et Z les résultats des dés de chaque couleur. On peut considérer la variable $X+Y+Z$ qui prend les valeurs de 3 à 18.

1.2 Linéarité de l'espérance et additivité de la variance

Théorème 1 : Soit X et Y deux variables aléatoires d'un univers Ω et $a \in \mathbb{R}$:

- **Linéarité de l'espérance** : $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ et $\mathrm{E}(aX) = a\,\mathrm{E}(X)$

Si les variables X et Y sont indépendantes :
- **Additivité de la variance** : $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ et $\mathrm{V}(aX) = a^2\mathrm{V}(X)$

Remarque : On considéra l'indépendance des variables au sens intuitif du terme c'est à dire que le résultat de X n'influe pas sur le résultat de Y comme dans le lancement de deux dés.

Exemple : De l'exemple précédent, on a :

- $\mathrm{E}(X) = \dfrac{1}{4}(1+2+3+4) = 2{,}5$ et $\mathrm{E}(Y) = \dfrac{1}{6}(1+2+3+4+5+6) = 3{,}5$.

  Donc $\mathrm{E}(X+Y) = 2{,}5 + 3{,}5 = 6$ et $\mathrm{E}(2X) = 2 \times 2{,}5 = 5$.

  La moyenne de la somme des résultats est 6 sur un grand nombre de lancers.

- $\mathrm{V}(X) = \dfrac{1}{4}(1+4+9+16) - 2{,}5^2 = 1{,}25$

  $\mathrm{V}(Y) = \dfrac{1}{6}(1+4+9+16+25+36) - 3{,}5^2 = \dfrac{35}{12} \approx 2{,}92$

  Donc $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) \approx 4{,}17$ et $\mathrm{V}(2X) = 4 \times 1{,}25 = 5$

Remarque : On peut généraliser ces résultats à la somme de n variables.

2 Somme de variables identiques et indépendantes

2.1 Décomposition d'une variable aléatoire suivant une loi binomiale

Théorème 2 : Soit n variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$ suivant la même loi de Bernoulli $\mathscr{B}(p)$.

La variable aléatoire $S_n = X_1 + X_2 + \cdots + X_n$ suit alors la loi binomiale $\mathscr{B}(n, p)$.

Démonstration : Soit les variables X_i suivant une même loi de Bernoulli $\mathscr{B}(p)$ qui prend la valeur 1 pour un succès avec $i \in [\![1, n]\!]$. Comme les variables X_i sont indépendantes, leur somme $X_1 + X_2 + \cdots + X_n$ prendra comme valeur le nombre de succès pour n expériences de Bernoulli, donc $S_n = X_1 + X_2 + \cdots + X_n$ suit la loi binomiale $\mathscr{B}(n, p)$.

Exemple : Soit X_i suivant une loi de Bernoulli $\mathscr{B}(0,13)$ pour $i \in [\![1, 10]\!]$, alors $S_{10} = X_1 + X_2 + \cdots + X_{10}$ suit la loi binomiale $\mathscr{B}(10\,;\,0,13)$.

Théorème 3 : Toute variable aléatoire X suivant la loi binomiale $\mathscr{B}(n, p)$ peut se décomposer en une somme de n variables indépendantes S_n.
$S_n = X_1 + X_2 + \ldots + X_n$ où X_i avec $i \in [\![1, n]\!]$ suit une même loi de Bernoulli $\mathscr{B}(p)$.

Remarque : Ce théorème permet de démontrer l'expression de l'espérance et de la variance d'une loi binomiale $\mathscr{B}(n, p)$.

En effet si X suit la loi binomiale $\mathscr{B}(n, p)$, on peut décomposer X en somme de n variables indépendantes suivant la loi de Bernoulli $\mathscr{B}(p)$ d'espérance p et de variance $p(1-p)$.

- $\mathrm{E}(X) = \mathrm{E}(X_1 + X_2 + \cdots + X_n) \overset{\text{linéarité}}{=} \underbrace{\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n)}_{n \text{ variables suivant } \mathscr{B}(p)} = np$
- $\mathrm{V}(X) = \mathrm{V}(X_1 + X_2 + \ldots + X_n) \overset{\text{additivité}}{=} \underbrace{\mathrm{V}(X_1) + \mathrm{V}(X_2) + \ldots + \mathrm{V}(X_n)}_{n \text{ variables indépendantes suivant } \mathscr{B}(p)} = np(1-p)$

Exemple : Soit X suivant la loi binomiale $\mathscr{B}(5\,;\,0,3)$ alors, on peut décomposer X en somme de 5 variables suivant la loi de Bernoulli $\mathscr{B}(0,3)$.

2.2 Échantillon d'une variable aléatoire

Définition 2 : Soit une variable X suivant une loi de probabilité.

Une liste de variables indépendantes $(X_1, X_2, \ldots, X_n)$ suivant cette même loi est appelée échantillon de taille n associé à X.

On pose $S_n = X_1 + X_2 + \cdots + X_n$ et $M_n = \dfrac{S_n}{n}$, on a alors :

$$\mathrm{E}(S_n) = n\,\mathrm{E}(X) \ , \ \mathrm{E}(M_n) = \mathrm{E}(X) \quad \text{et} \quad \mathrm{V}(S_n) = n\mathrm{V}(X) \ , \ \mathrm{V}(M_n) = \frac{\mathrm{V}(X)}{n}$$

Remarque : Plus la taille n de l'échantillon est grand plus la variance de M_n est petite donc plus la valeur de M_n se rapproche de l'espérance de X.

Exemple : Soit X une variable aléatoire dont la loi de probabilité est donné par le tableau suivant.

x_i	-10	5	20
$p(X = x_i)$	$\frac{1}{4}$	$\frac{11}{20}$	$\frac{1}{5}$

On considère un échantillon $(X_1, X_2, \dots, X_n)$ de la loi suivie par X et la variable aléatoire moyenne M_n

Déterminer la taille de l'échantillon n à partir de laquelle la variance de M_n devient inférieure à 0,05.

On calcule l'espérance et la variance de X :

- $\mathrm{E}(X) = \dfrac{1}{20}(-50 + 55 + 80) = \dfrac{85}{20} = \dfrac{17}{4}$
- $\mathrm{V}(X) = \dfrac{1}{20}(500 + 275 + 1600) - \left(\dfrac{17}{4}\right)^2 = \dfrac{475}{4} - \dfrac{289}{16} = \dfrac{1611}{16} \approx 100,7$

$\mathrm{V}(M_n) < 0,05 \quad \Leftrightarrow \quad \dfrac{\mathrm{V}(X)}{n} < 0,05 \quad \Leftrightarrow \quad n > \dfrac{\mathrm{V}(X)}{0,05} \approx 2\,014$

À partir d'un échantillon de 2 014 variables, la variance de M_n est inférieur à 0,05.

3 Concentration et loi des grands nombres

3.1 Inégalité de Bienaymé-Tchebychev

Théorème 4 : Soit X une variable aléatoire d'espérance μ et de variance V.

$$\forall \delta \in]0\,;+\infty[, \quad p\left(|X - \mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}}{\delta^2}$$

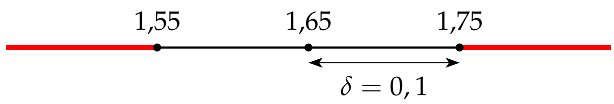
Remarque : La probabilité que X se trouve en dehors de l'intervalle $[\mu - \delta\,;\, \mu + \delta]$ est inférieur à $\dfrac{\mathrm{V}}{\delta^2}$. Cette inégalité conduit à la loi des grands nombres.

Exemple : La taille moyenne d'une femme française est de 1,65 m et la variance est évaluée à 0,002 5. Majorer la proportion des femmes françaises dont la taille est inférieure ou égale à 1,55 ou supérieure ou égale à 1,75.

Soit T_F la variable aléatoire associée à la taille d'une femme française. On a donc :

$$\mu = 1,65 \quad \text{et} \quad \mathrm{V} = 0,002\,5$$

$|T_F - 1,65| \geqslant 0,1$



On a alors : $p\left(|T_F - 1,65| \geqslant 0,1\right) \leqslant \dfrac{0,002\,5}{0,1^2} \quad \Leftrightarrow \quad p\left(|T_F - 1,65| \geqslant 0,1\right) \leqslant 0,25$

Il y a au plus un quart des femmes françaises dont la taille est inférieure ou égale à 1,55 ou supérieure ou égale à 1,75.

3.2 Application à un intervalle de rayon de k fois l'écart-type

Théorème 5 : Soit X une variable aléatoire d'espérance μ et d'écart-type σ.

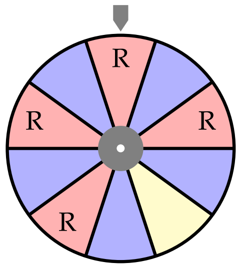
$$\forall k \in \mathbb{N}^*,\quad p\left(|X-\mu| \geqslant k\sigma\right) \leqslant \frac{1}{k^2}$$

Démonstration : On prend, avec $k \in \mathbb{N}^*$, $\quad \delta = k\sigma \quad \Rightarrow \quad \delta^2 = k^2\sigma^2 = k^2\mathrm{V}$

De inégalité de Bienaymé-Tchebychev :

$$p\left(|X-\mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}}{\delta^2} \overset{\delta=k\sigma}{\Rightarrow} p\left(|X-\mu| \geqslant k\sigma\right) \leqslant \frac{\mathrm{V}}{k^2\mathrm{V}} \;\Rightarrow\; p\left(|X-\mu| \geqslant k\sigma\right) \leqslant \frac{1}{k^2}$$

Exemple : Sur une roue de loterie il y a 4 secteurs rouges sur 10. On fait tourner 20 fois la roue en notant par X le nombre de fois où la roue tombe sur un secteur rouge.
La variable aléatoire X suit alors la loi binomiale $\mathscr{B}(20\,;\,0,4)$.
Majorer la probabilité que X soit en dehors de l'intervalle centrée en μ et de rayon 2σ.

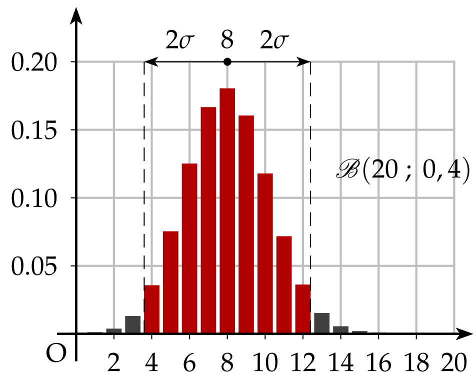


$\mu = 20\times 0,4 = 8$ et $\sigma = \sqrt{20\times 0,4\times 0,6} \approx 2,19$

On a alors : $p\left(|X-\mu| \geqslant 2\sigma\right) \leqslant \dfrac{1}{4}$

$2\sigma \approx 4,4$, à l'aide de la calculatrice, on trouve :

$$\begin{aligned} p\left(|X-8| \geqslant 4,4\right) &\overset{X\in\mathbb{N}}{=} p\left(|X-8| \geqslant 4\right) \\ &= p(X \leqslant 4) + p(X \geqslant 12) \\ &= p(X \leqslant 4) + 1 - p(X \leqslant 11) \\ &\approx 0,11 \end{aligned}$$



L'inégalité de Bienaymé-Tchebychev donne une majoration de 0,25 qui est loin d'être optimale comme le calcul sur la calculatrice le montre.

3.3 Inégalité de concentration

Théorème 6 : Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de variables aléatoires d'espérance μ et de variance V et M_n la variable aléatoire moyenne de cet échantillon.

$$\forall \delta \in]0\,;+\infty[,\quad p\left(|M_n-\mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}}{n\delta^2}$$

Remarque : On rappelle que $\quad M_n = \dfrac{X_1+X_2+\cdots+X_n}{n}$

Démonstration : D'après les relations sur l'espérance et la variance de la variable aléatoire d'un échantillon, on a $\quad \mathrm{E}(M_n) = \mu \quad$ et $\quad \mathrm{V}(M_n) = \dfrac{\mathrm{V}}{n}$.

D'après l'inégalité de Bienaymé-Tchebychev, on a :

$$p\left(|M_n-\mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}(M_n)}{\delta^2} \overset{\mathrm{V}(M_n)=\frac{\mathrm{V}}{n}}{\Leftrightarrow} p\left(|M_n-\mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}}{n\delta^2}$$

Exemple : On prend un dé tétraédrique bien équilibré dont on a déterminé au paragraphe 1.2 l'espérance $\mu = 2,5$ et la variance $\mathrm{V} = 1,25$.

Combien de lancers du dé tétraédrique doit-on faire pour s'assurer au seuil de 95 % que la moyenne des résultats des lancers est dans l'intervalle $]2,45\,;\,2,55[$.

Le rayon de l'intervalle $]2,45\,;\,2,55[$ est $\delta = 2,55 - 2,5 = 0,05$.

Soit M_n la variable aléatoire moyenne d'un échantillon de n lancers. On a alors :

$$p\left(|M_n - \mu| \geqslant \delta\right) \leqslant \frac{\mathrm{V}}{n\delta^2} \quad\Leftrightarrow\quad p\left(|M_n - 2,5| \geqslant 0,05\right) \leqslant \frac{1,25}{0,05^2\, n} \quad\Leftrightarrow$$

$$p\left(|M_n - 2,5| \geqslant 0,05\right) \leqslant \frac{500}{n} \quad\overset{\text{seuil de 95 \%}}{\Leftrightarrow}\quad \frac{500}{n} \leqslant 0,05 \quad\Leftrightarrow\quad n \geqslant \frac{500}{0,05} = 10\,000$$

Il faut faire au moins 10 000 lancers pour s'assurer que la moyenne des résultats est à moins de 5 % de l'espérance au seuil de 95 %.

Vérification : fonction simul() en Python .

```python
from random import*
def simul():
    s=0
    for i in range(10000):
        s=s+randint(1,4)
    return s/10000
```

En exécutant 4 fois simul(), on trouve :

2,4978 , 2,5073 , 2,4987 , 2,5149

Remarque : La majoration est loin d'être optimale, comme les valeurs trouvées, à moins de 2 % de l'espérance, le montrent !

3.4 Loi des grands nombres

Théorème 7 : Soit $(X_1, X_2, \dots, X_n)$ un échantillon de variables aléatoires d'espérance μ et M_n la variable aléatoire moyenne de cet échantillon.

$$\forall \delta \in]0\,;+\infty[,\quad \lim_{n\to+\infty} p\left(|M_n - \mu| \geqslant \delta\right) = 0$$

Remarque : Pour un δ donné aussi petit soit-il, la limite de la probabilité que M_n soit en dehors de l'intervalle $[\mu - \delta\,;\,\mu + \delta]$ est nul.

Ce théorème montre de façon rigoureuse, que lorsqu'on lance un grand nombre de fois une pièce de monnaie bien équilibrée, on a une chance sur deux en moyenne que la pièce tombe sur « pile » ou sur « face ».