Lycée Lucie Aubrac - Terminale

Sommes de variables aléatoires

Table des matières

1 Rappels sur les variables aléatoires **2**

2 Somme de variables aléatoires **2**

2.1 Définition 2

2.2 Espérance d'une somme 3

2.3 Variance d'une somme 4

3 Applications **7**

3.1 Application à la loi binomiale 7

3.2 Échantillon de n variables aléatoires indépendantes et identiquement distribuées 8

1 Rappels sur les variables aléatoires

Définition 1

Une **variable aléatoire** sur Ω est une fonction définie sur Ω à valeur dans $\mathbb{R}$.
À chaque issue de Ω, on associe un nombre réel.

Définition 2

Soit x un réel. L'événement « $\text{X} = x$ » est l'ensemble des issues de Ω auxquelles on associe le réel x, c'est-à-dire l'ensemble

$$\{\omega \in \Omega / X(\omega) = x\}.$$

Proposition 1

Si X est une variable aléatoire prenant les valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités $p_1, p_2, \ldots, p_n$ et si a et b des réels alors la loi de la variable aléatoire $Y = aX + b$ est donnée par le tableau suivant :

y_i	$ax_1 + b$	$ax_2 + b$	$\ldots$	$ax_n + b$
$P(Y = y_i)$	p_1	p_2	$\ldots$	p_n

Proposition 2

Si X est une variable aléatoire et a et b des réels alors

$$\text{E}(aX + b) = a\text{E}(X) + b.$$

Proposition 3

Si X est une variable aléatoire et a et b des réels alors

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

Proposition 4

Si X est une variable aléatoire et a et b des réels alors

$$\sigma(aX + b) = |a|\sigma(X).$$

2 Somme de variables aléatoires

2.1 Définition

Définition 3

Soient X et Y deux variables aléatoires définies sur l'univers Ω.
La somme de X et de Y est définie par la variable aléatoire Z telle que, pour tout $\omega \in \Omega$, $Z(\omega) = X(\omega) + Y(\omega)$.
On note $Z = X + Y$.

Exemple.
On lance deux dés à 6 faces. On note X le résultat du premier dé et Y le résultat du second dé. La variable aléatoire $Z = X + Y$ correspond à la somme des résultats des deux dés. Z prend ses valeurs dans l'ensemble $\{2\,;3\,;4\,;5\,;6\,;7\,;8\,;9\,;10\,;11\,;12\}$.

Proposition 5

Soient X et Y deux variables aléatoires définies sur l'univers Ω et à valeurs dans $\{0\,;1\,;\ldots;N\}$. Alors, pour tout $0 \leqslant n \leqslant N$:

$$\mathrm{P}(X+Y=n) = \sum_{k=0}^{n} \mathrm{P}\left(\{X=k\} \cap \{Y=n-k\}\right).$$

Démonstration.
Soit $0 \leqslant n \leqslant N$.

$$\begin{aligned}\{X+Y=n\} &= \Big(\{X=0\}\cap\{Y=n\}\Big)\bigcup\Big(\{X=1\}\cap\{Y=n-1\}\Big)\bigcup\ldots\bigcup\Big(\{X=n\}\cap\{Y=0\}\Big)\\ &= \bigcup_{k=0}^{n}\Big(\{X=k\}\cap\{Y=n-k\}\Big)\end{aligned}$$

Or, la réunion ci-dessus est une réunion disjointe d'événements. On obtient donc la formule souhaitée, à savoir :

$$\mathrm{P}(X+Y=n) = \sum_{k=0}^{n} \mathrm{P}\left(\{X=k\} \cap \{Y=n-k\}\right).$$

□

2.2 Espérance d'une somme

Proposition 6

Si X et Y sont deux variables aléatoires définies sur l'univers Ω, alors

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

Démonstration.
On note x_1, x_2, $\ldots$,x_n les valeurs prises par X et y_1, y_2, $\ldots$,y_m les valeurs prises par Y. Les valeurs prises par $X+Y$ sont x_1+y_1, x_1+y_2, $\ldots$($n \times m$ valeurs possibles au maximum). Les probabilités correspondantes sont $\mathrm{P}\left(\{X=x_1\}\cap\{Y=y_1\}\right)$, $\mathrm{P}\left(\{X=x_1\}\cap\{Y=y_2\}\right)$, $\ldots$

Par conséquent,

$$\begin{aligned}
\mathrm{E}(X+Y) &= \sum_{i,j}(x_i+y_j)\times \mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}x_i\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)+\sum_{i=1}^{n}\sum_{j=1}^{m}y_j\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}x_i\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)+\sum_{j=1}^{m}\sum_{i=1}^{m}y_j\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)\\
&= \sum_{i=1}^{n}x_i\sum_{j=1}^{m}\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)+\sum_{j=1}^{m}y_j\sum_{i=1}^{m}\mathrm{P}\left(\{X=x_i\}\cap\{Y=y_j\}\right)
\end{aligned}$$

Or, les événements $\{Y=y_1\},\{Y=y_2\},\ldots,\{Y=y_m\}$ forment une partition de l'univers. Ainsi, pour tout événement A, $\mathrm{P}(\mathrm{A})=\sum_{j=1}^{m}\mathrm{P}(\mathrm{A}\cap\{Y=y_i\})$
En particulier pour $\mathrm{A}=\{X=x_i\}$, on a, pour tout $1\leqslant i\leqslant n$:

$$\mathrm{P}(X=x_i)=\sum_{j=1}^{m}\mathrm{P}(\{X=x_i\}\cap\{Y=y_i\}).$$

De la même manière, pour tout $1\leqslant j\leqslant m$:

$$\mathrm{P}(Y=y_j)=\sum_{i=1}^{n}\mathrm{P}(\{X=x_i\}\cap\{Y=y_i\}).$$

On peut alors remplacer ces expressions dans l'expression de l'espérance de $X+Y$:

$$\begin{aligned}
\mathrm{E}(X+Y) &= \sum_{i=1}^{n}x_i\mathrm{P}(X=x_i)+\sum_{j=1}^{m}y_j\mathrm{P}(Y=y_j)\\
&= \mathrm{E}(X)+\mathrm{E}(Y)
\end{aligned}$$

□

Proposition 7

Si $X_1,X_2,\ldots,X_n$ sont des variables aléatoires définies sur l'univers Ω, alors

$$\mathrm{E}(X_1+X_2+\ldots+X_n)=\mathrm{E}(X_1)+\mathrm{E}(X_2)+\ldots+\mathrm{E}(X_n)$$

Démonstration.
Cette propriété découle de la Proposition 6 et se démontre par récurrence. □

2.3 Variance d'une somme

L'objectif de cette partie est de montrer que $V(X+Y)=V(X)+V(Y)$. Il faut toutefois préciser des conditions sur X et Y car cette formule n'est pas vraie en générale. Plus précisément, il faut que X et Y soient indépendantes pour que l'égalité soit vérifiée. La définition 4 donne ainsi un sens à l'indépendance de variables aléatoires. Les propositions suivantes (propositions 8 et 9) présentent ensuite deux lemmes préalables à la démonstration de l'égalité $V(X+Y) = V(X)+V(Y)$ pour des variables aléatoires indépendantes (proposition 10).

Définition 4

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires définies sur l'univers Ω et à valeur dans E. On dit qu'elles sont indépendantes si pour tous $x_1 \in E, x_2 \in E, \ldots, x_n \in E$,

$$\mathrm{P}(\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \ldots \cap \{X_n = x_n\}) = \mathrm{P}(X_1 = x_1) \times \mathrm{P}(X_2 = x_2) \times \ldots \times \mathrm{P}(X_n = x_n).$$

Remarque.
Dans le cas de deux variables aléatoires X et Y, il y a indépendance de X et Y si pour tout a et pour tout b, $\mathrm{P}(\{X = a\}cap\{Y = b\}) = \mathrm{P}(\{X = a\}) \times \mathrm{P}(\{Y = b\})$.
Autrement dit, X et Y sont indépendantes si, pour tout a et pour tout b, les événements $\{X = a\}$ et $\{Y = b\}$ sont indépendants.

Proposition 8 – (Lemme)

Si X et Y sont deux variables aléatoires indépendantes, alors

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y).$$

Démonstration.
On note x_1, x_2, $\ldots$,x_n les valeurs prises par X et y_1, y_2, $\ldots$,y_m les valeurs prises par Y. Les valeurs prises par $X+Y$ sont x_1y_1, x_1y_2, $\ldots$ ($n \times m$ valeurs possibles au maximum). Les probabilités correspondantes sont $\mathrm{P}\left(\{X = x_1\} \cap \{Y = y_1\}\right)$, $\mathrm{P}\left(\{X = x_1\} \cap \{Y = y_2\}\right)$, $\ldots$
Par conséquent,

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_{i,j} x_iy_j \times \mathrm{P}\left(\{X = x_i\} \cap \{Y = y_j\}\right) \\
&= \sum_{i=0}^{n} \sum_{j=0}^{m} x_iy_j \times \mathrm{P}\left(\{X = x_i\} \cap \{Y = y_j\}\right) \\
&= \sum_{i=0}^{n} \sum_{j=0}^{m} x_iy_j \times \mathrm{P}\left(\{X = x_i\}\right) \times \mathrm{P}\left(\{Y = y_j\}\right) \quad \text{(par indépendance de } X \text{ et } Y\text{)} \\
&= \sum_{i=0}^{n} x_i\mathrm{P}\left(\{X = x_i\}\right) \times \left(\sum_{j=0}^{m} y_j\mathrm{P}\left(\{Y = y_j\}\right)\right) \\
&= \sum_{i=0}^{n} x_i\mathrm{P}\left(\{X = x_i\}\right) \times \mathrm{E}(Y) \\
&= \mathrm{E}(Y) \times \left(\sum_{i=0}^{n} x_i\mathrm{P}\left(\{X = x_i\}\right)\right) \\
&= \mathrm{E}(Y) \times \mathrm{E}(X) \\
&= \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}$$

□

Proposition 9 – (Lemme) – Formule de König-Huygens

Soit $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire.

$$\mathrm{E}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 .$$

Démonstration.
On va écrire l'expression de la variance puis développer :

$$\begin{aligned}
V(X) =& \ p_1\,(x_1 - \mathrm{E}(X))^2 + p_2\,(x_2 - \mathrm{E}(X))^2 + \ldots + p_n\,(x_n - \mathrm{E}(X))^2 \\
=& \ p_1\left(x_1^2 - 2x_1\mathrm{E}(X) + (\mathrm{E}(X))^2\right) + p_2\left(x_2^2 - 2x_2\mathrm{E}(X) + (\mathrm{E}(X))^2\right) + \ldots + p_n\left(x_n^2 - 2x_n\mathrm{E}(X) + (\mathrm{E}(X))^2\right) \\
=& \ p_1x_1^2 + p_2x_2^2 + \ldots + p_nx_n^2 - 2\mathrm{E}(X)\,(p_1x_1 + p_2x_2 + \ldots + p_nx_n) + (\mathrm{E}(X))^2\,(p_1 + p_2 + \ldots + p_n)
\end{aligned}$$

Or, on sait que :
$p_1x_1 + p_2x_2 + \ldots + p_nx_n = \mathrm{E}(X)$;
$p_1 + p_2 + \ldots + p_n = 1$.
et $p_1x_1^2 + p_2x_2^2 + \ldots + p_nx_n^2 = \mathrm{E}(X^2)$;

En remplaçant dans l'expression, on obtient :

$$\begin{aligned}
V(X) &= \mathrm{E}(X^2) - 2\mathrm{E}(X) \times \mathrm{E}(X) + (\mathrm{E}(X))^2 \times 1 \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 .
\end{aligned}$$

□

Proposition 10

Si X et Y sont deux variables aléatoires indépendantes, alors

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

Démonstration.
Soient X et Y deux variables aléatoires indépendantes. D'après la proposition 8 (Formule de König-Huygens) :

$$\begin{aligned}
V(X+Y) &= \mathrm{E}\left((X+Y)^2\right) - \Big(\mathrm{E}(X+Y)\Big)^2 \\
&= \mathrm{E}\left(X^2 + 2XY + Y^2\right) - \Big(\mathrm{E}(X+Y)\Big)^2 \\
&= \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}(Y^2) - \Big(\mathrm{E}(X) + \mathrm{E}(Y)\Big)^2 \quad \text{(par linéarité de l'espérance)} \\
&= \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}(Y^2) - \Big(\mathrm{E}(X)\Big)^2 - 2\mathrm{E}(X)\mathrm{E}(Y) - \Big(\mathrm{E}(Y)\Big)^2 \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 + \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 + 2\mathrm{E}(XY) - 2\mathrm{E}(X)\mathrm{E}(Y) \\
&= V(X) + V(Y) \ \ \text{(en utilisant les propositions 8 et 9)}
\end{aligned}$$

□

Proposition 11

Si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires indépendantes définies sur l'univers Ω, alors

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$$

Démonstration.
Cette propriété découle de la Proposition 10 et se démontre par récurrence. □

3 Applications

3.1 Application à la loi binomiale

Définition 5

Deux variables aléatoires X et Y sont indentiquement distribuées si elles ont la même loi de probabilité.

Proposition 12 – (admise)

Toute variable aélatoire suivant une loi binomiale peut s'écrire comme la somme de variables aléatoires suivant une loi de Bernoulli indépendantes et indentiquement distribuées.

Remarque.
Si X_1, $\dots X_n$ sont des variables aléatoires indépendantes suivant une loi de Bernoulli (toutes de même paramètre p), la variable $X = X_1 + \dots + X_n$ suit une loi binomiale de paramètres n et p car elle compte le nombre de succès correspondant aux $(X_i)_{1\leqslant i\leqslant n}$. La proposition énonce le fait que toute variable aléatoire suivant une loi binomiale s'écrit de cette manière.

Proposition 13

Soit X une variable aléatoire suivant une loi binomiale.

- $\mathrm{E}(X) = np$
- $V(X) = np(1-p)$
- $\sigma(X) = \sqrt{np(1-p)}$

Démonstration.
Soit X une variable aléatoire suivant une loi binomiale. D'après la proposition **??**, on peut écrire $X = X_1 + \dots + X_n$ où $X_1, \dots, X_n$ sont des variables de Bernoulli indépendantes et identiquement distribuées. On a, pour tout $1 \leqslant i \leqslant n$, $\mathrm{E}(X_i) = p$ et $V(X_i) = p(1-p)$.
Ainsi,

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(X_1 + X_2 + \dots + X_n) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_2) + \dots + \mathrm{E}(X_n) \quad \text{(d'après la proposition 7)} \\
&= p + p + \dots + p \\
&= np
\end{aligned}$$

De plus,

$$\begin{aligned}
V(X) &= V(X_1 + X_2 + \dots + X_n) \\
&= V(X_1) + V(X_2) + \dots + V(X_n) \quad \text{(par indépendance des } X_i \text{ : proposition 11)} \\
&= p(1-p) + p(1-p) + \dots + p(1-p) \\
&= np(1-p)
\end{aligned}$$

Enfin,

$$\sigma(X) = \sqrt{V(X)} = \sqrt{np(1-p)}.$$

□

3.2 Échantillon de n variables aléatoires indépendantes et identiquement distribuées

Définition 6

Soit X une variable aléatoire. Un échantillon de taille n de X est la donnée de n variables aléatoires indépendantes $X_1, \ldots, X_n$ et qui sont toutes identiquement distribuées, suivant la même loi que X.

Proposition 14

Soient $X_1, X_2, \ldots, X_n$ des variables indépendantes et indentiquement distribuées. On note $S_n = X_1 + X_2 + \ldots + X_n$. Alors,

- $\mathrm{E}(S_n) = n\mathrm{E}(X_1)$
- $V(S_n) = nV(X_1)$
- $\sigma(S_n) = \sqrt{n}\sigma(X_1)$

Remarque.

- Cette proposition généralise le résultat de la partie précédente concernant la somme de variables de Bernoulli indépendantes et identiquement distribuées.
- X_1 ne joue pas un rôle particulier, on on pourrait remplacer X_1 par n'importe quel X_k (pour $1 \leqslant k \leqslant n$) dans les formules de la proposition 14.

Démonstration.

$$\begin{aligned}
\mathrm{E}(S_n) &= \mathrm{E}(X_1 + X_2 + \ldots + X_n) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_2) + \ldots + \mathrm{E}(X_n) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_1) \quad (\text{car pour tout } k,\ \mathrm{E}(X_k) = \mathrm{E}(X_1)) \\
&= n\mathrm{E}(X_1)
\end{aligned}$$

$$\begin{aligned}
V(S_n) &= V(X_1 + X_2 + \ldots + X_n) \\
&= V(X_1) + V(X_2) + \ldots + V(X_n)(\text{par indépendance des } X_i) \\
&= V(X_1) + V(X_1) + \ldots + V(X_1) \quad (\text{car pour tout } k,\ V(X_k) = V(X_1)) \\
&= nV(X_1)
\end{aligned}$$

□

Proposition 15

Soient $X_1, X_2, \ldots, X_n$ des variables indépendantes et indentiquement distribuées. On note $M_n = \dfrac{S_n}{n} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Alors,

- $\mathrm{E}(M_n) = \mathrm{E}(X_1)$
- $V(M_n) = \dfrac{V(X_1)}{n}$
- $\sigma(M_n) = \dfrac{\sigma(X_1)}{\sqrt{n}}$

Démonstration.

$$\begin{aligned}
\mathrm{E}(M_n) &= \mathrm{E}\left(\frac{S_n}{n}\right) \\
&= \frac{\mathrm{E}(S_n)}{n} \quad \text{(par linéarité de l'espérance)} \\
&= \frac{n\mathrm{E}(X_1)}{n} \quad \text{(d'après la proposition 14)} \\
&= \mathrm{E}(X_1)
\end{aligned}$$

$$\begin{aligned}
V(M_n) &= V\left(\frac{S_n}{n}\right) \\
&= \frac{\mathrm{E}(S_n)}{n^2} \\
&= \frac{nV(X_1)}{n^2} \quad \text{(d'après la proposition 14)} \\
&= \frac{V(X_1)}{n}
\end{aligned}$$

$$\begin{aligned}
\sigma(M_n) &= \sqrt{V\left(M_n\right)} \\
&= \sqrt{\frac{V(X_1)}{n}} \quad \text{(d'après l'expression de } V(M_n)\text{)} \\
&= \frac{\sqrt{V(X_1)}}{\sqrt{n}} \\
&= \frac{\sigma(X_1)}{\sqrt{n}}
\end{aligned}$$

□

Remarque.
L'égalité $\mathrm{E}(M_n) = \mathrm{E}(X_1)$ peut s'interpréter de la manière suivante : en prenant prenant un grand nombre d'échantillon de taille n, et en calculant la moyenne de chaque échantillon, la moyenne de ces moyennes sera environ $\mathrm{E}(X_1)$.

Exemple.
On lance cinq dés équilibrés et on note S la somme des dés. Déterminer $\mathrm{E}(S)$.

Solution :
La loi de probabilité de S est difficile à déterminer. Il suffit en revanche d'appliquer la proposition 14. Si on note X_1, X_2, X_3, X_4 et X_5 les résultats respectifs des cinq dés, elles sont indépendantes et indentiquement distribuées. De plus, $S = X_1 + X_2 + X_3 + X_4 + X_5$ et

$$\mathrm{E}(X_1) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = \frac{7}{2}.$$

Par conséquent,

$$\mathrm{E}(S) = 5 \times \mathrm{E}(X_1) = 5 \times \frac{7}{2} = \frac{35}{2}.$$

Exemple.
Supposons que dans le cadre d'un sondage pour un referendum, il y a une proportion p de la population qui souhaite voter Oui. On tire au hasard n personnes dans la population (ce tirage peut être modélisé par des variables aléatoires de Bernoulli $X_1, \dots X_n$ indépendantes et identiquement distribuées : elles sont toutes de paramètres p). Chaque X_i vaut 0 si la personne vote Non et 1 si la personne vote Oui. Déterminer $\mathrm{E}\left(\frac{X_1 + \ldots + X_n}{n}\right)$ et $\sigma\left(\frac{X_1 + \ldots + X_n}{n}\right)$ puis interpréter les résultats par une phrase.

Solution :
D'après la proposition 15 :

$$\mathrm{E}(M_n) = \mathrm{E}(X_1) = p \quad \text{et} \quad \sigma(M_n) = \frac{\sigma(X_1)}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}.$$

Interprétation : $M_n = \frac{X_1 + \ldots + X_n}{n}$ correspond à la proportion de personnes de l'échantillon ayant répondu Oui au sondage. De plus, les résultats obtenus signifient que si on prélève un grand nombre d'échantillons de taille n, la moyenne des proportions de ces échantillons sera environ p (la proportion de Oui dans la population générale) et que l'écart-type des proportions ces échantillons sera environ $\sqrt{\frac{p(1-p)}{n}}$.

Savoir-faire du chapitre

- Calculer l'espérance de la somme de variables aléatoires.
- Calculer la variance de la somme de variables aléatoires indépendantes.
- Représenter une variable comme somme de variables aléatoires plus simples, notamment pour modéliser un échantillonage.

QCM d'entrainement