Chapitre XII

La loi des grands nombres

I - Espérance et variance d'une variable aléatoire réelle

1 - Notion d'espérance mathématique

Définition : Espérance

Soit X une variable aléatoire réelle définie sur un univers Ω suivant la loi de probabilité suivante :

x_i	x_1	x_2	$\dots$	x_n
$P(X = x_i)$	p_1	p_2	$\dots$	p_n

L'**espérance** de X est le nombre réel noté $E(X)$ défini par :

$$E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n$$

Interprétation de l'espérance

$E(X)$ peut s'interpréter comme la moyenne des valeurs prises par X lorsque l'expérience aléatoire est répétée un **très grand nombre** de fois.

Exemple

X est la variable aléatoire qui suit la loi de probabilité suivante :

x_i	-2	1	4
$P(X = x_i)$	$0,2$	$0,5$	$0,3$

Son espérance est $E(X) = -2 \times 0,2 + 1 \times 0,5 + 4 \times 0,3 = 1,3$. Sur un très grand nombre de répétitions de cette expérience aléatoire, la valeur moyenne de X est $1,3$.

Propriété : Linéarité de l'espérance

Soient X une variable aléatoire et a et b des réels. Alors :

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

Exemple

On reprend l'exemple précédent et on définit la variable aléatoire Y par $Y = -2X + 10$. Son espérance est $E(Y) = E(-2X + 10) = -2E(X) + 10 = 7,4$.

2 - Notions de variance et d'écart-type

Définition : Variance

Soit X une variable aléatoire réelle définie sur un univers Ω suivant la loi de probabilités suivante :

x_i	x_1	x_2	$\dots$	x_n
$P(X = x_i)$	p_1	p_2	$\dots$	p_n

La **variance** de X est le nombre réel noté $V(X)$ défini par :

$$V(X) = p_1\left(x_1 - E(X)\right)^2 + \ldots + p_n\left(x_n - E(X)\right)^2.$$

Définition : Écart-type

En gardant les hypothèses et notations précédentes, l'**écart-type** de X est le nombre réel noté $\sigma(X)$ défini par :

$$\sigma(X) = \sqrt{V(X)}.$$

Interprétation

Plus la variance (et donc l'écart-type) sont grands, plus les valeurs prises par X sont dispersées autour de $E(X)$. Ce sont des indicateurs de **dispersion**.

Exemple

En gardant toujours le même exemple, on a $V(X) = 0,2(-2-1,3)^2 + 0,5(1-1,3)^2 + 0,3(4-1,3)^2 = 4,41$ et $\sigma = \sqrt{4,41} = 2,1$.

Propriété : La variance n'est pas linéaire

Soient X une variable aléatoire et a et b des réels. Alors :

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

Exemple

On reprend l'exemple précédent où $Y = -2X + 10$. Sa variance est $V(X) = (-2)^2 \times V(X) = 4 \times 4,41 = 17,64$.

3 - Dans le cadre d'une loi binomiale

Propriété : Espérance et variance d'une variable aléatoire suivant une loi binomiale

Soient n un entier naturel non nul et $p \in [0;1]$. Si X est une variable aléatoire suivant la loi binomiale de paramètres n et p alors

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

Exemple

Un sac contient 150 jetons indiscernables au toucher dont 105 sont gris et les autres sont marrons. On tire au hasard, successivement et avec remise, dix jetons de ce sac.

- On répète dix fois la même épreuve de Bernoulli de succès "obtenir un jeton gris" de paramètre $p = \frac{105}{150} = 0,7$ donc on a un schéma de Bernoulli de paramètres $n = 10$ et $p = 0,7$. La variable aléatoire X compte le nombre de succès et suit donc la loi binomiale de paramètres $n = 10$ et $p = 0,7$.
- Son espérance est $E(X) = 10 \times 0,7 = 7$.
- Interprétation : si l'on réitère un grand nombre de fois cette expérience, on obtient sept jetons gris en moyenne par expérience.

II - Somme de variables aléatoires

Définition : Somme de deux variables aléatoires

Si X et Y sont deux variables aléatoires, $X + Y$ est la variable aléatoire qui prend pour valeurs les sommes de valeurs possibles de X et de Y.

Exemple

X prend les valeurs -2, -1, 5 et 12 et Y prend les valeurs -5, -3, 0 et 1.

$X \mid Y$	-2	-1	5	12
-5	-7	-6	0	7
-3	-5	-4	2	9
0	-2	-1	5	12
1	-1	0	6	13

Alors les valeurs prises par la variable aléatoire $X + Y$ sont 7, -6, -5, -4, -1, 0, 1, 2, 5, 6, 7, 9, 12 et 13.

Propriété : Linéarité de l'espérance

Soient X et Y deux variables aléatoires. Alors :

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

Exemple 1

Soit X une variable aléatoire telle que $E(X) = 10$ et $V(X) = 4$. Soit Y une autre variable aléatoire telle que $E(Y) = 0$ et $V(Y) = 1$.

- Posons $Z = 2X - 15$. Alors par linéarité de l'espérance, $E(Z) = 2E(X) - 15 = 2 \times 10 - 15 = 5$ et $V(Z) = 2^2V(X) = 4 \times 4 = 16$.
- Posons $T = X + Y$. Alors par linéarité de l'espérance, $E(T) = E(X) + E(Y) = 10 + 0 = 10$. On ne sait pas encore calculer $V(X + Y)$, patientons...
- Posons $U = -5Y$. Alors $E(U) = -5E(Y) = -5 \times 0 = 0$ et $V(U) = (-5)^2V(Y) = 25 \times 1 = 25$.

Exemple 2

X est une variable aléatoire qui suit la loi binomiale de paramètres $n = 200$ et $p = 0,2$. Y est une variable aléatoire qui suit la loi binomiale de paramètres $n = 100$ et $p = 0,5$. Alors $E(X) = 200 \times 0,2 = 40$ et $E(Y) = 100 \times 0,5 = 50$.

$X + Y$ est une variable aléatoire dont les valeurs possibles sont $\{0; 1; 2; \ldots; 300\}$. D'après la linéarité de l'espérance, $E(X + Y) = E(X) + E(Y) = 40 + 50 = 90$.

Propriété : Variance de variables aléatoires indépendantes

Soient X et Y deux variables aléatoires associées à deux expériences dont les conditions de réalisation sont **indépendantes**. Alors

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

Exemple

X et Y sont deux variables aléatoires **indépendantes** telles que $V(X) = 1,25$ et $V(Y) = 5$. Donc $V(X + Y) = 1,25 + 5 = 6,25$.

Reprenons la variable aléatoire $T = X + Y$ de l'exemple 2 précédent. Rien n'indique que X et Y sont indépendantes. Ainsi **nous ne pouvons pas** conclure sur la valeur de $V(T)$.

III - La loi des grands nombres

Propriété : Inégalité de Bienaymé-Tchebychev

Soient X une variable aléatoire et δ un réel strictement positif. Alors

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

Interprétation

La probabilité que les valeurs prises par X s'éloignent de plus de δ de l'espérance diminue lorsque δ augmente.

Exemple : minorer ou majorer une probabilité

On lance 3600 fois une pièce de monnaie non truquée. Soit X la variable aléatoire égale au nombre de Piles obtenus. On cherche à minorer la probabilité que X soit compris entre 1600 et 2000.
X suit la loi binomiale de paramètres $n = 3600$ et $p = \frac{1}{2}$. Elle a pour espérance $E(X) = 3600 \times \frac{1}{2} = 1800$ et pour variance $V(X) = 3600 \times \frac{1}{2} \times \frac{1}{2} = 900$. On a :

$$\begin{aligned} P\big(1600 < X < 2000\big) &= P\big(-200 < X - 1800 < 200\big) \\ &= P\big(|X - 1800| < 200\big) \\ &= 1 - P\big(|X - 1800| \geq 200\big) \end{aligned}$$

D'après l'inégalité de Bienaymé-Tchebychev avec $\delta = 200$ on obtient :

$$P\big(|X - 1800| \geq 200\big) \leq \frac{900}{200^2}$$

D'où

$$\begin{aligned} 1 - P\big(|X - 1800| \geq 200\big) &\geq 1 - \frac{900}{200^2} \\ &\geq 1 - 0,0225 \\ &\geq 9,775 \end{aligned}$$

On a obtenu que $P\big(1600 < X < 2000\big) \geq 9,775$, autrement dit la probabilité d'obtenir strictement entre 1600 et 2000 Piles est au moins égale à 97,75 %.

Propriété : Inégalité de concentration

Soient $n \in \mathbb{N}$ et X_1, ..., X_n des variables aléatoires **indépendantes et de même loi de probabilité** admettant une espérance μ et une variance V.
Pour tout $n \in \mathbb{N}$, on pose M_n la variable aléatoire définie par $M_n = \frac{X_1 + \ldots + X_n}{n}$. Alors pour tout réel δ strictement positif on a :

$$P\big(|M_n - \mu| \geq \delta\big) \leq \frac{V}{n\delta^2}$$

Exemple : déterminer une taille d'échantillon

On effectue n tirages successifs, avec remise, d'une boule dans une urne contenant deux boules rouges et trois boules noires. On note X_i la variable variable aléatoire égale à 1 si la boule tirée au i-ème tirage est rouge et 0 sinon. À partir de combien de tirages peut-on être sûr à plus de 95 % que la proportion de boules rouges obtenues est strictement comprise entre 0,35 et 0,45 ?

Les variables aléatoires X_1, ..., X_n sont indépendantes et de même loi de Bernoulli de paramètre 0,4. Elles admettent donc pour espérance $\mu = 0,4$ et pour variance $V = 0,4 \times 0,6 = 0,24$.

On cherche le plus petit entier n tel que $P\big(0,35 < M_n < 0,45\big) \geq 0,95$, où $M_n = \dfrac{X_1 + \ldots + X_n}{n}$. On a :

$$\begin{aligned} P\big(0,35 < M_n < 0,45\big) &= P\big(-0,05 < M_n - 0,4 < 0,05\big) \\ &= P\big(|M_n - 0,4| < 0,05\big) \\ &= 1 - P\big(|M_n - 0,4| \geq 0,05\big) \end{aligned}$$

D'après l'inégalité de concentration avec $\delta = 0,05$ on obtient :

$$P\big(|M_n - 0,4| \geq 0,05\big) \leq \frac{0,24}{n \times 0,05^2}$$

D'où

$$\begin{aligned} 1 - P\big(|M_n - 0,4| \geq 0,05\big) &\geq 1 - \frac{0,24}{n \times 0,05^2} \\ P\big(0,35 < M_n < 0,45\big) &\geq 1 - \frac{96}{n} \end{aligned}$$

On cherche le plus petit entier n vérifiant $1 - \dfrac{96}{n} \geq 0,95$. La résolution de l'inéquation nous donne finalement $n = 1920$.

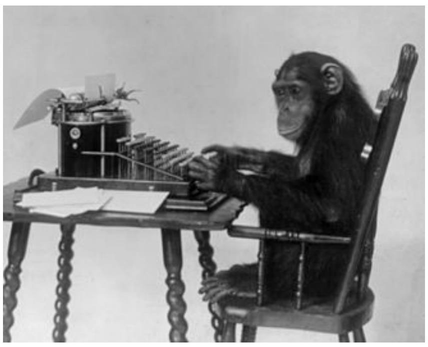
Propriété : Loi des grands nombres

Avec les mêmes hypothèses et notations que précédemment, on a :

$$\lim_{n \to +\infty} P\big(|M_n - \mu| \geq \delta\big) = 0$$

Interprétation

Plus n est grand, plus la moyenne empirique M_n est proche de l'espérance μ.



Ce chimpanzé ne tapera probablement pas Les Misérables, sauf si on lui laisse du temps.