LA LOI DES GRANDS NOMBRES

I. Somme de variables aléatoires :

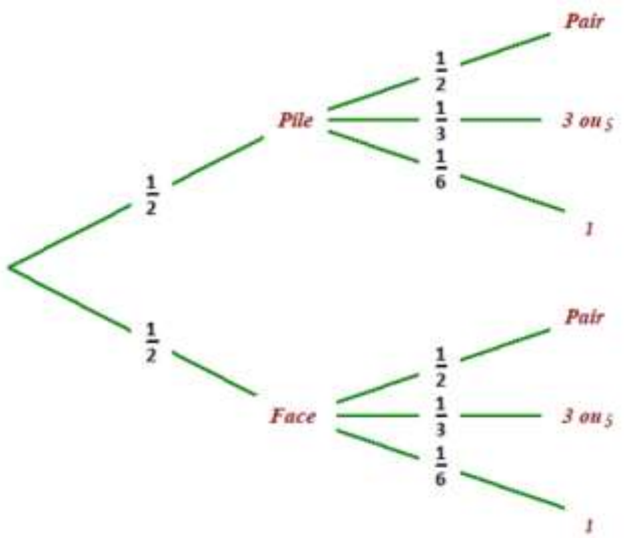
1) Exemple :

On considère le jeu suivant qui se déroule en deux parties.

1ère partie : On lance une pièce de monnaie bien équilibrée.
Si on tombe sur "pile", on gagne 1€.
Si on tombe sur "face" on gagne 2€.

2è partie : On lance un dé à 6 faces, non pipé.
Si on tombe sur un chiffre pair, on gagne 1€.
Si on tombe sur le 3 ou le 5, on gagne 2€.
Si on tombe sur le 1, on perd 5€.

La variable aléatoire X désigne le gain obtenu à l'issue de la première partie.
La variable aléatoire Y désigne le gain obtenu à l'issue de la deuxième partie.
On peut considérer ces deux variables aléatoires comme indépendantes.
Etablir la loi de probabilité de la variable aléatoire S = X + Y donnant le gain à la fin des deux parties.
(on pourra s'aider d'un arbre probabilisé)



Valeurs de X	Valeurs de Y	Valeurs de S
1€	1€	2€
1€	2€	3€
1€	– 5€	– 4€
2€	1€	3€
2€	2€	4€
2€	– 5€	– 3€

$S = s_i$	$P(S = s_i)$
– 4€	$P(S = -4) = P(X = 1) \times P(Y = -5) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$
– 3€	$P(S = -3) = P(X = 2) \times P(Y = -5) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$
2€	$P(S = 2) = P(X = 1) \times P(Y = 1) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
3€	$P(S = 3) = P(X = 1) \times P(Y = 2) + P(X = 2) \times P(Y = 1) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$
4€	$P(S = 4) = P(X = 2) \times P(Y = 2) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$
TOTAL	$\frac{1}{12} + \frac{1}{12} + \frac{1}{4} + \frac{5}{12} + \frac{1}{6} = 1$

2) Définitions :

Soient X et Y deux variables aléatoires.

La loi de probabilité de la variable aléatoire somme S = X + Y est donnée par :

i,j et k étant trois réels

$$P(X+Y=k) = \sum_{i+j=k} P(\,(X=i) \cap (Y=j))$$

Si X et Y sont indépendantes,

$$P(X+Y=k) = \sum_{i+j=k} P(X=i) \times P(Y=j)$$

3) Propriétés de l'espérance et de la variance :

a et b réels

E(aX + b) = a E(X) + b

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

V(aX + b) = a² V(X)

V(X + Y) = V(X) + V(Y)   si X et Y sont indépendantes.

Exemple :

Une entreprise qui fabrique des roulements à bille fait une étude sur une gamme de billes produites.
Le diamètre théorique doit être égal à 1,3 cm mais cette mesure peut être légèrement erronée.
L'expérience consiste à tirer au hasard une bille d'un lot de la production et à mesurer son diamètre.
On considère la variable aléatoire X qui, à une bille choisie au hasard, associe son diamètre.
La loi de probabilité de X est résumée dans le tableau suivant :

x_i	1,298	1,299	1,3	1,301	1,302
$P(X=x_i)$	0,2	0,1	0,2	0,4	0,1

1) Calculer l'espérance et l'écart-type de la variable aléatoire X.

$E(X) = 1{,}298 \times 0{,}2 + 1{,}299 \times 0{,}1 + 1{,}3 \times 0{,}2 + 1{,}301 \times 0{,}4 + 1{,}302 \times 0{,}1 = 1{,}3001$

$\sigma(X) = \sqrt{1{,}298^2 \times 0{,}2 + 1{,}299^2 \times 0{,}1 + 1{,}3^2 \times 0{,}2 + 1{,}301^2 \times 0{,}4 + 1{,}302^2 \times 0{,}1 - 1{,}3001^2} = 0{,}0013$

2) Posons Y = 1000X – 1300.

Calculer l'espérance, la variance et l'écart-type de Y.

y_i	– 2	– 1	0	1	2
$P(Y = y_i)$	0,2	0,1	0,2	0,4	0,1

$E(Y) = -2 \times 0{,}2 - 1 \times 0{,}1 + 0 \times 0{,}2 + 1 \times 0{,}4 + 2 \times 0{,}1 = 0{,}1$

$\sigma(Y) = \sqrt{(-2)^2 \times 0{,}2 + (-1)^2 \times 0{,}1 + 0^2 \times 0{,}2 + 1^2 \times 0{,}4 + 2^2 \times 0{,}1 - 0{,}1^2} = 1{,}3$

En déduire l'espérance, la variance et l'écart-type de X.

$E(Y) = 1000\, E(X) - 1300 \Leftrightarrow E(X) = \dfrac{E(Y) + 1300}{1000} = \dfrac{0{,}1 + 1300}{1000} = 1{,}3001$

$V(Y) = 1000^2\, V(X) \Leftrightarrow V(X) = \dfrac{V(Y)}{1000^2}$ donc $\sigma(X) = \sqrt{\dfrac{V(Y)}{1000^2}} = \dfrac{\sigma(Y)}{1000} = 0{,}0013$

4) <u>Application à la loi binomiale</u> :

Soit X_1 , X_2 , … , X_n n variables aléatoires indépendantes suivant la même loi.
On dit que l'on a un échantillon de variables aléatoires, de taille n.

Posons $S = X_1 + X_2 + \ldots + X_n$.

Alors $\mathbf{E(S) = E(X_1) + E(X_2) + \ldots + E(X_n)}$ **et** $\mathbf{V(S) = V(X_1) + V(X_2) + \ldots + V(X_n)}$

<u>Cas particulier</u> : si X_1 , X_2 , … , X_n suivent une loi de Bernoulli de paramètre p
c'est-à-dire que $P(X_i = 1) = p$ et $P(X_i = 0) = 1 - p$.
S suit une loi binomiale de paramètres n et p.
$E(S) = E(X_1) + E(X_2) + \ldots + E(X_n) = p + p + \ldots + p = np$
et
$V(S) = V(X_1) + V(X_2) + \ldots + V(X_n) = p(1 - p) + p(1 - p) + \ldots + p(1 - p) = np(1 - p)$

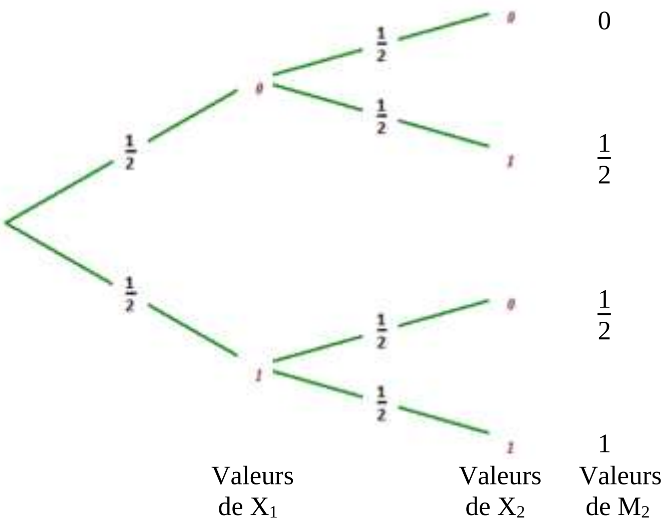
II. Concentration, loi des grands nombres :

1) Moyenne d'un échantillon de n variables aléatoires :

a) <u>Exemple</u> :

On lance un dé à 6 faces.
Si on obtient un nombre pair, X prend la valeur 1
Si on obtient un nombre impair, X prend la valeur 0.

1) On répète 2 fois l'expérience et on s'intéresse à la valeur moyenne obtenue.
On pose X_1 la variable aléatoire associée au 1^er^ lancé
et X_2 la variable aléatoire associée au 2^è^ lancé.
(X_1 , X_2) est alors un échantillon de 2 variables aléatoires suivant chacune une loi de Bernoulli de paramètre $\frac{1}{2}$. On pose M_2 la variable aléatoire moyenne de X_1 et X_2 .
Déterminer les différentes valeurs prises par M_2 puis calculer l'espérance et la variance de M_2.

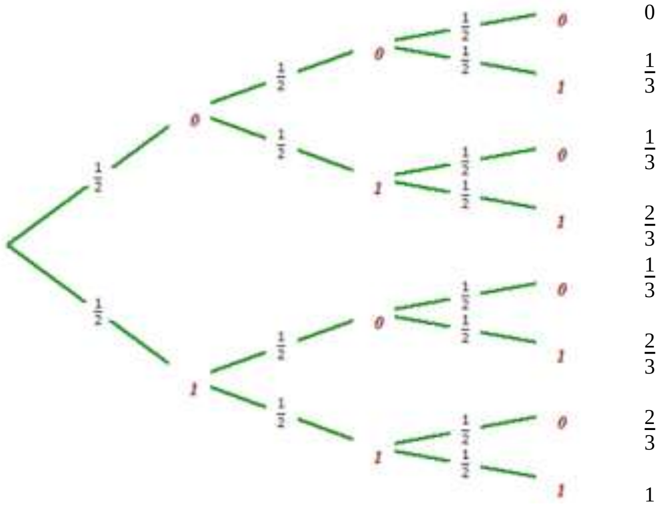


m_i	0	$\frac{1}{2}$	1	Total
$P(M_2 = m_i)$	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \times 2 = \frac{1}{2}$	$\frac{1}{4}$	1

$E(M_2) = 0 \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} + 1 \times \frac{1}{4} = \frac{1}{2}$ et $E(X_1) = E(X_2) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2} = E(M_2)$

$V(M_2) = 0^2 \times \frac{1}{4} + \left(\frac{1}{2}\right)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{4} - \frac{1}{4} = \frac{1}{8}$ et $V(X_1) = V(X_2) = 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4} = 2\,V(M_2)$

2) On répète maintenant 3 fois l'expérience. Recommencer l'exercice pour M_3 la variable aléatoire moyenne de X_1, X_2 et X_3.



Valeurs de X_1 | Valeurs de X_2 | Valeurs de X_3 | Valeurs de M_3

m_i	0	$\frac{1}{3}$	$\frac{2}{3}$	1	Total
$P(M_2 = m_i)$	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$	$\frac{1}{8} \times 3 = \frac{3}{8}$	$\frac{1}{8} \times 3 = \frac{3}{8}$	$\frac{1}{8}$	1

$E(M_3) = 0 \times \frac{1}{8} + \frac{1}{3} \times \frac{3}{8} + \frac{2}{3} \times \frac{3}{8} + 1 \times \frac{1}{8} = \frac{1}{2}$ et $E(X_1) = E(X_2) = E(X_3) = \frac{1}{2} = E(M_3)$

$V(M_3) = 0^2 \times \frac{1}{8} + \left(\frac{1}{3}\right)^2 \times \frac{3}{8} + \left(\frac{2}{3}\right)^2 \times \frac{3}{8} + 1^2 \times \frac{1}{8} - \frac{1}{4} = \frac{1}{12}$ et $V(X_1) = V(X_2) = V(X_3) = \frac{1}{4} = 3\, V(M_3)$

b) <u>Définition</u> :

Soit (X_1 , X_2 , … , X_n) un échantillon de n variables aléatoires indépendantes suivant la même loi.

Soit M_n la variable aléatoire moyenne de l'échantillon définie par $M_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

On a : $\mathbf{E(M_n) = E(X_1) = E(X_2) = \ldots = E(X_n)}$

$$\mathbf{V(M_n) = \frac{1}{n} V(X_1) = \frac{1}{n} V(X_2) = \ldots = \frac{1}{n} V(X_n)}$$

$$\boldsymbol{\sigma(M_n) = \frac{1}{\sqrt{n}} \sigma(X_1) = \frac{1}{\sqrt{n}} \sigma(X_2) = \ldots = \frac{1}{\sqrt{n}} \sigma(X_n)}$$

2) Inégalité de Markov :

a) Propriété :

Si X est une variable aléatoire réelle positive ou nulle d'espérance E(X), alors, pour tout réel a strictement positif, on a : $\mathbf{P(\ X \geq a\ ) \leq \dfrac{E(X)}{a}}$

b) Application :

Sur une autoroute, la vitesse moyenne des automobilistes est de 120km.h^{-1}.

1) Majorer la probabilité qu'un automobiliste roule à une vitesse supérieure à 150 km.h^{-1}.

Posons X la variable aléatoire qui représente la vitesse d'un automobiliste en km.h^{-1}.
X est de façon évidente positive et E(X) = 120.

Alors, d'après l'inégalité de Markov, on a $P(\ X \geq 150\ ) \leq \dfrac{120}{150}$ donc $P(\ X \geq 150\ ) \leq \dfrac{4}{5}$

2) Minorer la probabilité qu'un automobiliste roule à une vitesse inférieure à 100 km.h^{-1}.

D'après l'inégalité de Markov, on a $P(\ X \geq 100\ ) \leq \dfrac{120}{100}$ donc $P(\ X \geq 100\ ) \leq \dfrac{6}{5}$

$P(\ X < 100\ ) = 1 - P(\ X \geq 100\ )$

$P(\ X \geq 100\ ) \leq \dfrac{6}{5} \Leftrightarrow -P(\ X \geq 100\ ) \geq -\dfrac{6}{5} \Leftrightarrow 1 - P(\ X \geq 100\ ) \geq 1 - \dfrac{6}{5}$

$\Leftrightarrow \quad P(\ X < 100\ ) \geq -\dfrac{1}{5}$

L'inégalité de Markov ne donne donc dans ce cas aucune indication intéressante !

3) Inégalité de Bienaymé-Tchébychev :

a) Propriété :

Si X est une variable aléatoire, pour tout réel strictement positif δ on a :

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

b) Application :

X est une variable aléatoire suivant une loi binomiale de paramètres n = 20 et p = 0,1.

1) Appliquer l' inégalité de Bienaymé-Tchébychev avec $\delta = 2\sigma(X)$.

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

$P(\ |\ X - E(X)\ | \geq 2\sigma(X)\ ) \leq \dfrac{1}{4}$ mais $\sigma(X) = \sqrt{np(1-p)} = \sqrt{20 \times 0{,}1 \times 0{,}9} = \sqrt{0{,}18}$

$E(X) = np = 20 \times 0{,}1 = 2$

$P(\ |\ X - 2\ | \geq 2\sqrt{0{,}18}\ ) \leq \dfrac{1}{4}$

2) Recommencer avec $\delta = 3\sigma(X)$ puis avec $\delta = 4\sigma(X)$.

$P(\ |\ X - 2\ | \geq 3\sqrt{0{,}18}\ ) \leq \dfrac{1}{9}$

$P(\ |\ X - 2\ | \geq 4\sqrt{0{,}18}\ ) \leq \dfrac{1}{16}$

3) Que constate-t-on ?

On constate que les écarts entre X et son espérance supérieurs à quelques σ ont une probabilité de plus en plus faible de se produire.

4) **Inégalité de concentration :**

a) Propriété :

Soit (X_1 , X_2 , … , X_n) un échantillon de n variables aléatoires indépendantes suivant la même loi qu'une variable aléatoire X.
Soit M_n la variable aléatoire moyenne de l'échantillon.
Pour tout réel strictement positif δ on a :

$$\mathbf{P(| M_n - E(\mathit{X}) | \geq \delta) \leq \frac{V(\mathit{X})}{n\,\delta^2}}$$

b) Exemple :

On peut se servir de l'inégalité de concentration pour déterminer la taille d'un échantillon.

X est une variable aléatoire suivant une loi de Bernoulli de paramètre p = 0,2.
On a un échantillon de n variables aléatoires indépendantes suivant la même loi de probabilité que *X* et de variable aléatoire moyenne Mn .
Déterminer la taille n de l'échantillon telle que $P(M_n \in [0,03 ; 0,37]) \geq 0,95$.

Le centre de l'intervalle [0,03 ; 0,37] est $\frac{0,03 + 0,37}{2} = 0,2$.

La longueur de l'intervalle est 0,37 – 0,03 = 0,34 donc le rayon de l'intervalle est $\frac{0,34}{2} = 0,17$.

Dire que $M_n \in [0,03 ; 0,37]$ signifie que $d(M_n ; 0,2) \leq 0,17$ donc $| M_n - 0,2 | \leq 0,17$.
X est une variable aléatoire suivant une loi de Bernoulli de paramètre p = 0,2 donc E(*X*) = 0,2.
On a donc $P(| M_n - 0,2 | \leq 0,17) \geq 0,95$ donc $P(| M_n - E(X) | \leq 0,17) \geq 0,95$
Pour se rapprocher de l'inégalité de concentration, on va prendre l'événement contraire
donc $P(| M_n - 0,2 | \geq 0,17) \leq 0,05$
Posons alors δ = 0,17. V(*X*) = p (1 – p) = 0,2 × 0,8 = 0,16.
Si on a $\frac{V(X)}{n\,\delta^2} \leq 0,05$ l'inégalité sera vérifiée.

Donc il faut chercher n tel que $\frac{V(X)}{n\,\delta^2} \leq 0,05 \Leftrightarrow V(X) \leq n\,\delta^2 \times 0,05 \Leftrightarrow n \geq \frac{V(X)}{0,05\,\delta^2}$

$$\Leftrightarrow n \geq \frac{0,16}{0,05 \times 0,17^2}$$

$$\Leftrightarrow n \geq 110,7$$

A partir de 111 variables aléatoires dans l'échantillon, la probabilité que la variable aléatoire moyenne soit comprise entre 0,03 et 0,17 est supérieure à 0,95.

5) Loi faible des grands nombres :

a) Propriété :

Soit un échantillon de n variables aléatoires indépendantes suivant la même loi de probabilité qu'une variable aléatoire *X* et de variable aléatoire moyenne Mn.
Pour tout réel strictement positif δ on a :

$$\mathbf{\lim_{n \to +\infty} P(| M_n - E(\mathit{X}) | \geq \delta) = 0}$$

On peut également l'exprimer ainsi :
Plus la taille de l'échantillon est importante, plus la probabilité que l'écart entre la moyenne de l'échantillon et l'espérance de *X* dépasse δ est faible donc plus cet écart est faible.