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)

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.
2) Posons Y = 1000X – 1300.
   Calculer l'espérance, la variance et l'écart-type de Y.
   En déduire l'espérance, la variance et l'écart-type de X.

4) Application à la loi binomiale :

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

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

Cas particulier : si $X_1, X_2, \dots, X_n$ 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.
$E(S) = E(X_1) + E(X_2) + \dots + E(X_n) = p + p + \dots + p = np$
et
$V(S) = V(X_1) + V(X_2) + \dots + V(X_n) = p(1-p) + p(1-p) + \dots + p(1-p) = np(1-p)$

II. Concentration, loi des grands nombres :

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

a) Exemple :

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 1er lancé
et X_2 la variable aléatoire associée au 2è 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.
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.

b) Définition :

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

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

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

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

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

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}.

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

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 $\delta = 2\sigma(X)$.

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

3) Que constate-t-on ?

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 :

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

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 $P(M_n \in [0{,}03 ; 0{,}37]) \geq 0{,}95$.

5) Loi 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 :

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

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.