Sommes de Variables Aléatoires – Tale spé maths

A) Étude d'une variable aléatoire (Rappels)

1) étude d'un exemple

Exercice : On donne les 2 jeux suivants associés aux variables aléatoires X et Y :

Jeu N°1

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

Quel est le jeu le plus intéressant ?

Jeu N°2

y_i	-3	-1	0	1	2
$P(Y=x_i)$	0,1	0,4	0,2	0,2	0,1

Justifier la réponse avec des calculs appropriés

2) Espérance mathématique

Définition : Soit X une variable aléatoire tel que $X(\Omega)=\{1;2;3;\dots;n\}$;

l'espérance mathématique de X est le réel $\boldsymbol{E(X)=\sum_{i=1}^{n} x_i . P(X=x_i)}$

Application : Dans le jeu n°1 on obtient

$E(X)=(-5)\times 0,2+(-1)\times 0,3+0\times 0,1+1\times 0,1+3\times 0,3=\mathbf{-0,3}$

Dans le jeu n°2 on obtient

$E(Y)=(-3)\times 0,1+(-1)\times 0,4+0\times 0,2+1\times 0,2+2\times 0,1=\mathbf{-0,3}$

3) Variance & écart-type

Définition : Soit X une variable aléatoire tel que $X(\Omega)=\{1;2;3;\dots;n\}$;

- la variance de X est le réel positif $\boldsymbol{V(X)=\sum_{i=1}^{n}(x_i-E(X))^2 . P(X=x_i)}$
- l'écart-type de X est le réel positif : $\boldsymbol{\sigma(X)=\sqrt{V(X)}}$

Application : Dans le jeu n°1 on obtient

$V(X)=(-5+0,3)^2\times 0,2+(-1+0,3)^2\times 0,3+(0+0,3)^2\times 0,1+\cdots$

$\cdots+(1+0,3)^2\times 0,1+(3+0,3)^2\times 0,3=\mathbf{8,01}$

donc on déduit que : $\sigma(X)=\sqrt{8,01}\simeq\mathbf{2,83}$

Dans le jeu n°2 on obtient

$V(Y)=(-3+0,3)^2\times 0,2+(-1+0,3)^2\times 0,4+(0+0,3)^2\times 0,2+\cdots$

$\cdots+(1+0,3)^2\times 0,2+(2+0,3)^2\times 0,1=\mathbf{1,81}$

donc on déduit que : $\sigma(X)=\sqrt{1,81}\simeq 1,35$

<u>*Conclusion*</u> : le jeu n°1 paraît plus risqué que le jeu n°2

B) Étude de deux variables aléatoires

1) Analyse d'un exemple

Exercice : On pose X et Y les variables aléatoires définies par les lois suivantes

x_i	-4	1	20
$P(X=x_i)$	0,1	0,35	0,55

y_i	-2	5
$P(Y=y_i)$	0,27	0,73

a) Soit Z la variable aléatoire définie par $Z=X+Y$. Quelles sont les valeurs prises par la variable aléatoire Z ?

b) Peut-on déterminer la loi de probabilité de Z à partir des données de l'énoncé ? Si oui, donner cette loi

solution : on a $X(\Omega)=\{-4;1;20\}$ et $Y(\Omega)=\{-2;5\}$

	-4	1	20
-2	-6	-1	18
5	1	6	25

les valeurs possibles de Z sont

$Z(\Omega)=\{-6;-1;1;6;18;25\}$ (cf tableau ci-contre)

les probabilités liées à ces valeurs de Z sont :

$p(Z=-6)=p((X=-4)\cap(Y=-2))=0,1\times 0,27=0{,}027$

$p(Z=-1)=p((X=1)\cap(Y=-2))=0,35\times 0,27=0{,}0945$

$p(Z=1)=p((X=-4)\cap(Y=5))=0,1\times 0,73=0{,}073$

$p(Z=6)=p((X=1)\cap(Y=5))=0,35\times 0,73=0{,}2555$

$p(Z=18)=p((X=20)\cap(Y=-2))=0,55\times 0,27=0{,}1485$

$p(Z=25)=p((X=20)\cap(Y=5))=0,55\times 0,73=0{,}4015$

rque : on vérifie la validité de la loi de Z avec $\sum_{k\in Z(\Omega)} k.P(Z=k)=1$

Problème : que se passe-t-il si plusieurs valeurs de Z correspondent à des cas distincts de X et de Y ?

2) Étude de la variable aléatoire Z=X+Y

Définition : Soit X et Y deux variables aléatoires définies sur un univers fini Ω d'expérience aléatoire, qui prennent respectivement pour valeurs les réels x_i tels que $1\leqslant i\leqslant n$ et y_j tels que $1\leqslant j\leqslant m$, avec n et m entiers naturels.
Alors si on pose $Z=X+Y$, on a :

- $Z(\Omega)=\{x_i+y_j\ /\ 1\leqslant i\leqslant n,\ 1\leqslant j\leqslant m\}=\{k_i/0\leqslant i\leqslant N\}$ où $N=n+m$
- $\forall k\in Z(\Omega)$, $P(Z=k)=P(X+Y=x_i+y_j)=P(\{k_i\})$

Théorème : Soit la variable aléatoire $Z=X+Y$ où X et Y suivent des lois distinctes alors l'espérance mathématique de Z est $\mathbf{E(X+Y)=E(X)+E(Y)}$

Preuve :

Soient X, Y et Z trois variables aléatoires définies sur Ω tel que $Z=X+Y$.

On a alors $E(X+Y)=E(Z)=\sum\limits_{i=0}^{n} Z(\omega_i)P(\{\omega_i\})=\sum\limits_{i=0}^{n}(X+Y)(\omega_i)P(\{\omega_i\})$.

On a par ailleurs $(X+Y)(\omega_i)=X(\omega_i)+Y(\omega_i)$.

Donc $E(X+Y)=\sum\limits_{i=0}^{n}(X)(\omega_i)P(\{\omega_i\})+\sum\limits_{i=0}^{n}Y(\omega_i)P(\{\omega_i\})$.

D'où $E(X+Y)=E(X)+E(Y)$ en identifiant les deux sommes précédentes à $E(X)$ et $E(Y)$.

Remarque : on peut généraliser ce théorème avec le ***théorème de transfert*** *« hors-programme »*:si f est continue alors : $\mathbf{E(f(X))=\sum_{i=1}^{n} f(x_i).P(\{x_i\})}$

Définition : Soit X et Y deux variables aléatoires définies sur un univers fini Ω on appelle la covariance du couple (X,Y) le réel issue de la *formule Huygens* :
$cov(X,Y)=E(X.Y)-E(X).E(Y)$

Propriété : Soit la variable aléatoire $Z=X+Y$ où X et Y suivent des lois distinctes alors la variance de Z est $V(Z)=V(X)+V(Y)+2cov(X,Y)$

Preuve : (hors-programme) on applique la *formule de Koënig* :
Pour toute variable aléatoire X : $V(X)=E(X^2)-E^2(X)$

Exercice : Calculer $E(X+Y)$ dans l'étude en B)1)
[réponse : E(X+Y)=14,06 ; E(X)=10,95 ; E(Y)=3,11]

3) Étude de la transformation affine Y=aX+b

Propriété : Soit X une variable aléatoire définie sur un univers fini Ω ; soit Y la variable obtenue par transformation affine $Y=aX+b$; alors on a :

$$\mathbf{E(Y)=a.E(X)+b}\ ;\quad \mathbf{V(Y)=a^2.V(X)}\ ;\quad \boldsymbol{\sigma(Y)=|a|.\sigma(X)}$$

Preuve : on a par définition $E(X)=\sum_{i=1}^{n} x_i.P(X=x_i)$; soient $a,b\in\mathbb{R}$

donc $a.E(X)=\sum_{i=1}^{n}(a.x_i).P(X=x_i)$ or $P(X=x_i)=P(aX=a.x_i)$

donc $a.E(X)=\sum_{i=1}^{n}(a.x_i).P(aX=a.x_i)$ de même $P(X=x_i)=P(X+b=x_i+b)$

donc $a.E(X)+b=\sum_{i=1}^{n}(a.x_i).P(aX=a.x_i)+b$ or $\sum_{i=1}^{n}P(X=x_i)=1$

donc $a.E(X)+b=\sum_{i=1}^{n}(a.x_i).P(aX=a.x_i)+b.\sum_{i=1}^{n}P(X=x_i)$

donc $a.E(X)+b=\sum_{i=1}^{n}(a.x_i).P(aX+b=a.x_i+b)+b.\sum_{i=1}^{n}P(aX+b=a.x_i+b)$

donc $a.E(X)+b=\sum_{i=1}^{n}(a.x_i+b).P(aX+b=a.x_i+b)$

donc $a.E(X)+b=\sum_{i=1}^{n}(y_i).P(Y=y_i)=E(Y)$

Exemple : On joue à un jeu, dans une foire, se déroulant en deux étapes.

- Dans la **phase 1**, on lance un dé équilibré à six faces. Si le résultat obtenu est 1 ou 6, on gagne 9 points. Sinon, on perd 6 points
- Dans la **phase 2**, on lance une pièce équilibrée. Si on obtient face,on gagne 6 points. Sinon, on perd 2 points.

Soit X la variable aléatoire du total de points obtenus. Calculons $E(X)$. Soient X_1 la variable aléatoire du gain obtenu à la 1ère étape et X_2 la variable aléatoire du gain obtenu à la 2nde étape. Donc on obtient $X=X_1+X_2$;

On étudie les lois de X_1 et X_2 ci-dessous

x_i	-6	9
$P(X_1=x_i)$	$\frac{2}{3}$	$\frac{1}{3}$

x_i	-2	6
$P(X_2=x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

ainsi $E(X_1)=-1$ et $E(X_2)=2$

donc $E(X)=E(X_1)+E(X_2)=-1+2=1$

C) Variables aléatoires indépendantes

1) Propriété de la variance

Propriété : Soient X et Y deux variables aléatoires indépendantes ; alors on a : $\boldsymbol{V(X+Y)=V(X)+V(Y)}$ ou encore $\boldsymbol{cov(X,Y)=0}$

Preuve : on sait que X et Y sont indépendantes donc $E(X.Y)=E(X)\times E(Y)$
donc $cov(X,Y)=E(X.Y)-E(X).E(Y)=0$ donc $V(X+Y)=V(X)+V(Y)$

Exercice : Calculer $V(X+Y)$ puis $\sigma(X+Y)$ dans le jeu de la « foire »
on a $V(X)=36\times\frac{2}{3}+81\times\frac{1}{3}-(-1)^2=50$ et $V(Y)=4\times\frac{1}{2}+36\times\frac{1}{2}-(2)^2=16$
donc $V(X+Y)=66$ donc $\sigma(X+Y)=\sqrt{66}\simeq 8{,}12$

2) Somme d'un échantillon

Propriété : On considère un entier naturel $n\geqslant 1$ et $X_1,\ldots X_n$, n variables aléatoires définies sur Ω supposées indépendantes et identiquement distribuées.
On note la **somme** de l'échantillon : $S_n=X_1+X_2+\cdots+X_n$ alors :

$$\boldsymbol{E(S_n)=n.E(X)} \ , \ \boldsymbol{V(S_n)=n.V(X)} \ , \ \boldsymbol{\sigma(S_n)=\sqrt{n}.\sigma(X)}$$

Preuve : $E(S_n)=E(X_1+\ldots+X_n)=E(X_1)+\ldots+E(X_n)=n.E(X)$
$V(S_n)=V(X_1+\ldots+X_n)=V(X_1)+\ldots+V(X_n)=n.V(X)$
donc $\sigma(S_n)=\sqrt{n.V(X)}=\sqrt{n}.\sqrt{V(X)}=\sqrt{n}.\sigma(X)$

Exercice : Dans le jeu de « *yam's* » on lance 5 dés et on effectue la somme des points obtenus ; on a $E(X)=3{,}5$; $V(X)\simeq 2{,}92$ et $\sigma(X)\simeq 1{,}71$
donc $E(S_5)=5\times 3{,}5=17{,}5$; $V(S_5)=5\times 2{,}92=14{,}6$ et
$\sigma(S_5)=\sqrt{5}\times 1{,}71=3{,}82$; obtient donc $17{,}5\,pts\pm 3{,}82\,pts$

3) Moyenne d'un échantillon

Propriété : On considère un entier naturel $n\geqslant 1$ et $X_1,\ldots X_n$, n variables aléatoires définies sur Ω supposées indépendantes et identiquement distribuées.
On note la **moyenne** de l'échantillon : $M_n=\dfrac{X_1+X_2+\cdots+X_n}{n}$ alors :

$$\boldsymbol{E(M_n)=E(X)} \ , \ \boldsymbol{V(M_n)=\frac{V(X)}{n}} \ , \ \boldsymbol{\sigma(M_n)=\frac{\sigma(X)}{\sqrt{n}}}$$

Preuve : $E(M_n)=E(\frac{1}{n}S_n)=\frac{1}{n}E(S_n)=\frac{1}{n}\times nE(X)=E(X)$

$V(M_n)=V(\frac{1}{n}S_n)=\frac{1}{n^2}V(S_n)=\frac{1}{n^2}\times nV(X)=\frac{V(X)}{n}$

et donc : $\sigma(M_n)=\sqrt{\dfrac{V(X)}{n}}=\dfrac{\sigma(X)}{\sqrt{n}}$ pour tout $n\in\mathbb{N}$

Exercice : On propose un jeu de gain d'argent de « tickets à gratter » suivant :

a	0	5	10	20	50	100
$P(X=a)$	0,6	0,2	0,1	0,06	0,03	0,01

On souhaite déterminer la moyenne des gains obtenus pour 10 tickets
on a $E(X)=5{,}7$, $V(X)\simeq 181{,}44$ et $\sigma(X)\simeq 13{,}47$
donc $E(M_{10})=5{,}7$; $V(M_{10})=\dfrac{181{,}44}{10}=18{,}144$, $\sigma(M_{10})=\dfrac{13{,}47}{\sqrt{10}}\simeq 4{,}26$

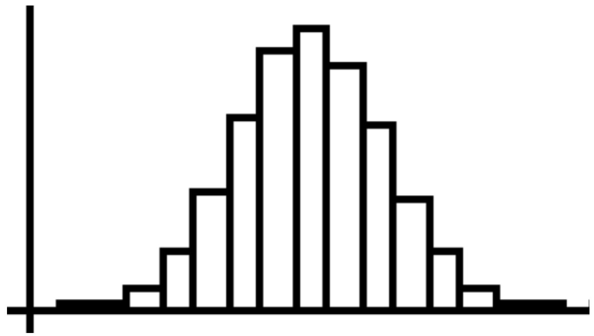
4) Application à la loi Binomiale

Propriété : Soient $X_1,\ldots,X_n$ des variables aléatoires suivantes toutes la même loi de Bernoulli $B(p)$ avec $p\in[0;1]$ on pose : $S_n=X_1+X_2+\cdots+X_n$
alors on obtient :

- la variable S_n suit la loi Binomiale $B(n,p)$
- $E(S_n)=np$; $V(S_n)=np(1-p)$; $\sigma(S_n)=\sqrt{np(1-p)}$

Exercice : On étudie la somme de 20 variables aléatoires de *Bernoulli* avec $p=0{,}4$; ainsi la variable aléatoire « somme » suit la loi *Binomiale* $B(20;0{,}4)$; $E(S)=8$; $V(S)=4{,}8$ et $\sigma(S)=2{,}19$

on obtient le diagramme de la distribution de S :



On souhaite déterminer des *intervalles de fluctuation* de cette variable :

- $P(5{,}8\leqslant S\leqslant 10{,}2)$
- $P(3{,}6\leqslant S\leqslant 12{,}4)$
- $P(1{,}4\leqslant S\leqslant 14{,}6)$

Comment calculer précisément ces 3 valeurs ?

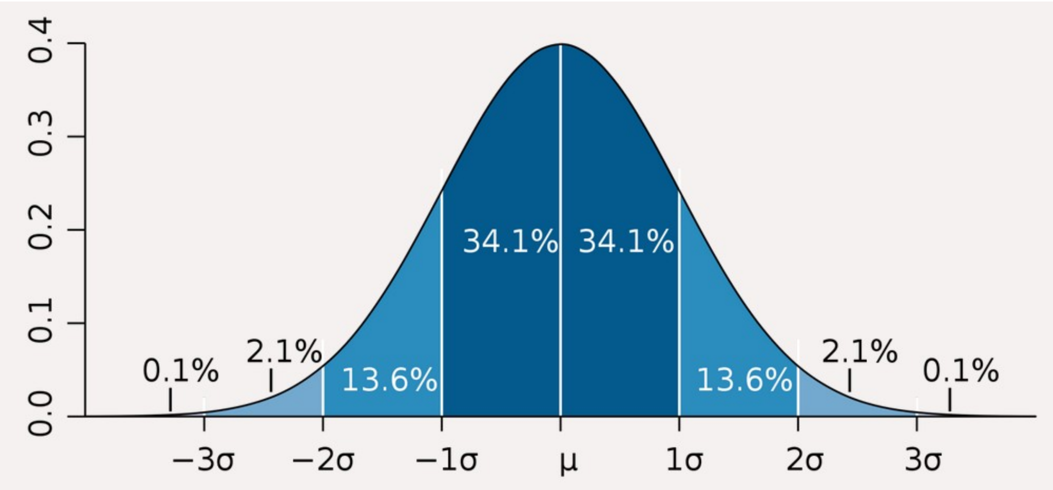
D) Concentration & loi des grands nombres

1) Les intervalles de Normalité

Théorème de Laplace-Gauss : On considère une variable aléatoire X suivant la loi Binomiale $B(n,p)$ avec $n\geqslant 30$ et $p\in[0;1]$ avec $np\geqslant 5$ et $n(1-p)\geqslant 5$; alors si on pose $E(X)=\mu=np$ et $\sigma(X)=\sigma=\sqrt{np(1-p)}$:

- $P(\mu-\sigma\leqslant X\leqslant\mu+\sigma)=0{,}6826\simeq\mathbf{0{,}68}$
- $P(\mu-2\sigma\leqslant X\leqslant\mu+2\sigma)=0{,}9544\simeq\mathbf{0{,}95}$
- $P(\mu-3\sigma\leqslant X\leqslant\mu+3\sigma)=0{,}9974\simeq\mathbf{0{,}99}$

<u>Rque</u> : Ces valeurs sont obtenues par la courbe de GAUSS : $f(x)=\frac{1}{\sqrt{2\pi}}.e^{-0{,}5x^2}$



on a : $\int_{-1}^{1} f(x).dx\simeq 0{,}6826$; $\int_{-2}^{2} f(x).dx\simeq 0{,}9544$; $\int_{-3}^{3} f(x).dx\simeq 0{,}9974$

<u>Rque²</u> : On dit alors que X suit une « *Loi Normale* » $N(\mu;\sigma)$

2) Inégalité de Markov

Propriété : Si X est une variable aléatoire discrète à valeurs positives et soit a un réel strictement positif, alors : $\boldsymbol{P(X\geqslant a)\leqslant\frac{E(X)}{a}}$

Preuve : Soit $A=\{x\in X(\Omega)/x\geqslant a\}$ et $B=\{x\in X(\Omega)/x<a\}$ alors $A\cap B=\emptyset$ et $A\cup B=X(\Omega)$ donc l'ensemble { A , B } forme une partition de l'Univers donc on obtient les inégalités suivantes ;

$$E(X)=\sum_{k\in X(\Omega)} k.P(X=k)=\sum_{k\in A} k.P(X=k)+\sum_{k\in B} k.P(X=k)$$
$$\geqslant\sum_{k\in A} k.P(X=k)=\sum_{k\geqslant a} k.P(X=k)\geqslant\sum_{k\geqslant a} a.P(X=k)$$
$$=a.\sum_{k\geqslant a} P(X=k)=a.P(X\geqslant a) \quad \text{donc} \quad aP(X\geqslant a)\leqslant E(X) \ [QED]$$

Propriété : Si X est une variable aléatoire discrète à valeurs positives et soit a un réel strictement positif, alors : $\boldsymbol{P(|X|\geqslant a)\leqslant\frac{E(X^2)}{a^2}}$

Preuve : on applique l'inégalité de *Markov* : $P(X^2\geqslant a^2)\leqslant\frac{E(X^2)}{a^2}$

or $X^2\geqslant a^2$ équivaut à $|X|\geqslant a$ donc on déduit : $P(|X|\geqslant a)\leqslant\frac{E(X^2)}{a^2}$

Exercice : Le nombre de pièces sortant d'une usine en une journée est une variable aléatoire d'espérance 50. On veut estimer la probabilité que la production d'un jour donné dépasse 75 pièces.

On a : $P(X\geqslant 75)\leqslant\frac{E(X)}{75}$ donc $P(X\geqslant 75)\leqslant\frac{50}{75}=\frac{2}{3}$ soit environ 67 %

En fait cette probabilité est trop grande et ne donne pas assez d'informations

3) Inégalité de *Bienaymé-Tchebychev*

Propriété : Si X est une variable aléatoire discrète à valeurs positives et soit δ un réel strictement positif, alors : $\boldsymbol{P(|X-E(X)|\geqslant\delta)\leqslant\frac{V(X)}{\delta^2}}$

Preuve : on applique la 2nde inégalité de *Markov* :

$P(|X-E(X)|\geqslant a)\leqslant\frac{E((X-E(X))^2)}{a^2}$ et $(X-E(X))^2=X^2-2X.E(X)+E^2(X)$

donc $E((X-E(X))^2)=E(X^2)-E^2(X)=V(X)$ donc $P(|X-E(X)|\geqslant\delta)\leqslant\frac{V(X)}{\delta^2}$

Exercice : On reprend l'exercice précédent ; Que peut-on dire de plus sur cette probabilité si on sait que la variance de la production quotidienne est 25?

$P(X\geqslant 75)\leqslant P(|X-50|\geqslant 25)\leqslant\frac{V(X)}{25^2}$ donc $P(X\geqslant 75)\leqslant\frac{25}{625}=\frac{1}{25}=0{,}04$

Soit une estimation plus précise de 4 %

Cela est bien sûr logique puisque l'on connaît maintenant 2 paramètres sur X

Exercice : Soit une variable aléatoire X qui suit la loi binomiale de paramètres $n=20$ et $p=0,4$; on souhaite comparer les valeurs de normalités de Gauss on sait que : $E(X)=8$; $V(X)=4,8$; $\sigma(X)\simeq 2,2$

si $\delta=\sigma(X)=2,2$ alors $P(|X-8|\geqslant 2,2)\leqslant\dfrac{4,8}{2,2^2}=0,99$

si $\delta=2\sigma(X)=4,4$ alors $P(|X-8|\geqslant 4,4)\leqslant\dfrac{4,8}{4,4^2}=0,25$

si $\delta=3\sigma(X)=6,6$ alors $P(|X-8|\geqslant 6,6)\leqslant\dfrac{4,8}{6,6^2}=0,11$

si $\delta=4\sigma(X)=8,8$ alors $P(|X-8|\geqslant 8,8)\leqslant\dfrac{4,8}{8,8^2}=0,06$

<u>*Conclusion*</u> ; l'inégalité de *Bienaymé-Tchebychev* est très intéressante pour observer les valeurs "extrêmes" c'est-à-dire très éloignées de la moyenne

<u>*Rque*</u> : On peut modéliser le problème à l'aide d'un script en langage PYTHON

Soit une variable aléatoire X qui suit la loi binomiale de paramètres $n=20$ et $p=0,1$.

```python
import random as rd
import math
def simulX():
  a=0
  for expe in range (20):
    if rd.randint(1,100)<=10:
      a=a+1
  return a
def proba(N):
  echant=[simulX() for i in range (N)]
  c=0
  d=2*math.sqrt(1.8)
  for e in echant:
    if abs(e-2)>=d:
      c=c+1
  return c/N
```

on a : $E(X)=2$; $V(X)=1,8$ et $\sigma(X)=\sqrt{1,8}\simeq 1,34$

On constate qu'un écart à $E(X)$ supérieur à $2\sigma(X)$ est de probabilité souvent inférieur à 0, 05 alors que l'inégalité de *Bienaymé-Tchebytchev* nous donne pour cette même probabilité une majoration par 0,25.

$$P(|X-2|\geqslant 2\sqrt{1,8})\leqslant\frac{1,8}{4\times 1,8}=0,25$$

```
>>>>proba(1000)
0.038
>>>>proba(10000)
0.0454
>>>>proba(100000)
0.041278
>>>>proba(1000000)
0.04516
```

l'inégalité est donc loin d'être optimale.

4) Inégalité de concentration

Propriété : Soit une variable aléatoire moyenne M_n d'un échantillon de taille n de la variable aléatoire X . Pour tout réel strictement positif δ , on a

$$\boldsymbol{P(|M_n-E(X)|\geqslant\delta)\leqslant\frac{V(X)}{n.\delta^2}}$$

Exercice : Soit une variable aléatoire X qui suit la loi de Bernoulli de paramètre $p=0,2$. On considère un échantillon de n variablesaléatoires suivant la loi de X. On appelle M_n la variable aléatoire moyenne associée à cet échantillon. Déterminer la taille n de l'échantillon pour que la condition $P(0,03<M_n<0,37)\geqslant 0,95$ soit satisfaite ; on rappelle que $E(M_n)=E(X)=0,2$

la condition donne : $P(|M_n-0,2|>0,17)\leqslant 0,05$ donc $\dfrac{V(X)}{n.\delta^2}\leqslant 0,05$ avec $V(X)=p(1-p)=0,16$ et $\delta=0,17$ donc $n\geqslant 110,73$ soit $n\geqslant\mathbf{111}$

5) Loi des grands nombres

Propriété : Soit une variable aléatoire moyenne M_n d'un échantillon de taille n de la variable aléatoire X . Pour tout réel strictement positif δ , on a

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

<u>*Rque*</u> : La loi des grands nombres traduit le fait que plus la taille de l'échantillon d'une variable aléatoire X est grande, plus l'écart entre la moyenne de cet échantillon et l'espérance de la variable aléatoire X est faible

Exercice : On considère la variable aléatoire X qui prend ses valeurs de manière équiprobable parmi les entiers 1 à 5. On nomme M_n la variable aléatoire moyenne d'un échantillon de taille n de la variable aléatoire X

Éditer un script PYTHON afin de simuler et vérifier la "*loi des grands nombres*"