❉ Chapitre 17 ❉

Loi des grands nombres

I. Rappel : Moyenne d'un échantillon

Exemple 1:

On lance un dé à six faces et on considère la variable aléatoire X qui prend la valeur 1 si le dé s'arrête sur un chiffre pair et la valeur 0 sinon.

X suit donc une loi de Bernoulli de paramètre $\frac{1}{2}$.

On répète deux fois de suite cette expérience. On considère alors l'échantillon $(X_1; X_2)$ de taille 2 de variables aléatoires X_1 et X_2 suivant la même loi que X. Il est ainsi possible d'évaluer le résultat d'une telle expérience en étudiant la variable aléatoire moyenne de X_1 et X_2.

On appelle M_2 la variable aléatoire moyenne de l'échantillon $(X_1; X_2)$. Alors M_2 peut prendre les valeurs suivantes :

Valeur de X_1	0	0	1	1
Probabilités de X_1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Valeur de X_2	0	1	0	1
Probabilités de X_2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Probabilités de (X_1, X_2)	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$
Valeur de M_2	$\frac{0+0}{2}=0$	$\frac{1+0}{2}=\frac{1}{2}$	$\frac{1+0}{2}=\frac{1}{2}$	$\frac{1+1}{2}=1$
Probabilités de M_2	$\frac{1}{4}$	$\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$		$\frac{1}{4}$

On obtient ainsi la loi de probabilité de M_2 :

x_i	0	$\frac{1}{2}$	1
$P(M_2 = x_i$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Définition 1:

Soient $(X_1, X_2, ..., X_n)$ un échantillon de taille n de variables aléatoires indépendantes suivant une meme loi.
La variable aléatoire moyenne M_n de l'échantillon est donnée par :

$$M_n = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

Propriété 1 :

Soit une variable aléatoire X et soit un échantillon $(X_1, X_2, ..., X_n)$ de taille n de varaibles aléatoires indépendantes suivant une même loi que X.

$$E(M_n) = E(X) \qquad V(M_n) = \frac{1}{n}V(x) \qquad \sigma(M_n) = \frac{1}{\sqrt{n}}\sigma(X)$$

II. Inégalité de Bienaymé-Tchebytchev

Propriété 2 : *Admise*

Soit une variable aléatoire X. Pour tout réel strictement positif δ, on a :

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

Méthode 1 :

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

1. Appliquer l'inégalité de Bienaymé-Tchebytchev avec $\delta = 2\sigma(X)$. Interpréter.
2. Recommencer avec $\delta = 2\sigma(X)$ et $\delta = 2\sigma(X)$. Que constate-t-on ?

1. $E(X) = 20 \times 0,1 = 2$ $\quad V(X) = 20 \times 0,1 \times 0,9 = 1,8$ $\quad \sigma(X) = \sqrt{1,8}$

$$P\left(|X-E(X)| \geqslant 2\sigma(X)\right) \leqslant \frac{V(X)}{(2\sigma(X))^2} \quad \text{donc} \quad P\left(|X-2)| \geqslant 2\sqrt{1,8}\right) \leqslant 0,25$$

La probabilité que l'écart de X à $E(X)$ soit supérieur à $2\sigma(X)$ est majorée par $0,25$.

2.
   - pour $\delta = 3\sigma(X)$: $P\left(|X-E(X)| \geqslant 3\sigma(X)\right) \leqslant \dfrac{V(X)}{(3\sigma(X))^2}$ donc $P\left(|X-2)| \geqslant 3\sqrt{1,8}\right) \leqslant \dfrac{1}{9}$
   - pour $\delta = 4\sigma(X)$: $P\left(|X-E(X)| \geqslant 4\sigma(X)\right) \leqslant \dfrac{V(X)}{(4\sigma(X))^2}$ donc $P\left(|X-2)| \geqslant 4\sqrt{1,8}\right) \leqslant 0,0625$
   - On peut en déduire que les écarts de X à $E(X)$ de quelques σ deviennent improbables.

Remarque :

On peut modéliser le problème à l'aide un programme informatique :

```
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
```

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

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 $0,25$. l'inégalité est donc loin d'être optimale.

III. Inégalité de concentration

Propriété 3 : *Admise*

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 :

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

Méthode 2 :

Soit une variable aléatoire X qui suit la loi de Bernoulli de paramètre $0,2$. On considère un échantillon de n variables aléatoires suivant la loi de X. On appelle M_n la variable aléatoire moyenne associée à cet échantillon. Déterminer la taille n de l'échantillon tel que la probabilité que la moyenne M_n appartienne à l'intervalle $]0,03;0,37]$ soit supérieur à $0,95$.

On cherche à calculer n tel que $P(0,03 < M_n < 0,37) \geqslant 0,95$. Dans l'idée d'appliquer l'inégalité de concentration, on fait apparaitre l'espérance de X dans l'inégalité. Or $E(X) = p = 0,2$.
Ainsi, on cherche n tel que :

$$\begin{aligned}
P(0,03-0,2 < M_n - 0,2 < 0,37-0,2) &\geqslant 0,95\\
P(-0,17 < M_n - 0,2 < 0,17) &\geqslant 0,95\\
P(|M_n - 0,2| < 0,17) &\geqslant 0,95\\
1 - P(|M_n - 0,2| > 0,17) &\geqslant 0,95\\
P(|M_n - 0,2| > 0,17) &\leqslant 0,05
\end{aligned}$$

En prenant $\delta = 0,17$ dans l'inégalité de concentration, on a : $P\big(|M_n - E(X)| \geqslant \delta\big) \leqslant 0,05$, avec $\dfrac{V(X)}{n\delta^2} = 0,05$.
Or, $V(X) = p(1-p) = 0,2 \times 0,8 = 0,16$
On cherche un entier n tel que : $\dfrac{0,16}{n0,17^2} \leqslant 0,05 \iff n \geqslant 110,7$
Pour $n \geqslant 111$, la probabilité que la moyenne M_n appartienne à l'intervalle $]0,03;0,37[$ est supérieur à $0,95$.

IV. Loi des grands nombres

Propriété 4 : *Admise*

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 :

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

Remarque :

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.

Exemple 2:

Simulons des valeurs d'une variable aléatoire moyenne dans le but d'observer la loi des grands nombres. 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.

$E(X) = 3$, $\quad V(X) = 2$, $\quad \sigma(X) = \dfrac{\sqrt{2}}{\sqrt{n}}$, $\quad$ donc $P\big(|M_n - E(X)| \geqslant \sigma(M_n)\big) = P\left(|M_n - 3| \geqslant \dfrac{\sqrt{2}}{\sqrt{n}}\right)$

```python
import random as rd
import math
def simulMn(n):
  S=[rd.randint(1,5) for i in range(n)]
  Mn=sum(S)/n
  return Mn
def echantMn(n):
  echant=[simulMn(n) for i in range (500)]
  c=0
  d=math.sqrt(2/n)
  for e in echant:
    if abs(e-3)>=d:
      c=c+1
  return c/500
```

```
>>>>echantMn(5)
0.31
>>>>echantMn(10)
0.036
>>>>echantMn(20)
0.008
>>>>echantMn(50)
0.0
```

Plus la taille de l'échantillon est grand, plus l'écart entre la moyenne de cet échantillon et l'espérance de la variable aléatoire X est faible.