

Cours – Séquence n° 18 – PROBABILITÉS (III) : concentration, loi des grands nombres

I- INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV



Irénée-Jules Bienaymé (1796-1878) est un mathématicien français. Il contribue au développement de la statistique et à ses applications à différents domaines (calculs financiers, démographie,...). Il traduit les travaux de son ami russe Tchebychev.



Pafnouti Tchebychev (1821-1894) est un mathématicien russe. Connu pour ses travaux en probabilité, il poursuit les recherches de Bernoulli et Poisson. Il démontre des théorèmes limites comme la loi des grands nombres.

Extrait de la revue Tangente N°197 (Décembre 2020)

Dans le cadre d'un article pour défendre la méthode des moindres carrés de Laplace face aux critiques de Denis Poisson, Bienaymé énonce et démontre, en 1853, l'inégalité qui porte leurs deux noms. Tchebychev se rend compte de l'importance de ce résultat passé inaperçu par son ami français. Il la publie et l'utilise pour démontrer la loi des grands nombres dans un cadre général. La notoriété du Russe la fait connaître mais la véritable inégalité est sans doute qu'elle ne porte que le nom de Tchebychev, excepté dans la littérature française.

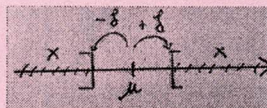
Propriété (admise) : inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance μ et de variance V .

Pour tout réel $\delta > 0$, $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$.

Remarques

- $|X - \mu| \geq \delta$ signifie que $X \notin]\mu - \delta ; \mu + \delta[$.
- Donc $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$ peut aussi s'écrire $P(X \notin]\mu - \delta ; \mu + \delta[) \leq \frac{V}{\delta^2}$.
- On peut déduire que $P(|X - \mu| < \delta) = P(X \in]\mu - \delta ; \mu + \delta[) \geq 1 - \frac{V}{\delta^2}$.
- Si $\delta \leq \sigma$ où σ est l'écart type de X , alors $\delta^2 \leq \sigma^2$ et donc $\frac{V}{\delta^2} \geq \frac{V}{\sigma^2}$ soit $\frac{V}{\delta^2} \geq 1$.



Dans ce cas particulier, l'inégalité est évidente.

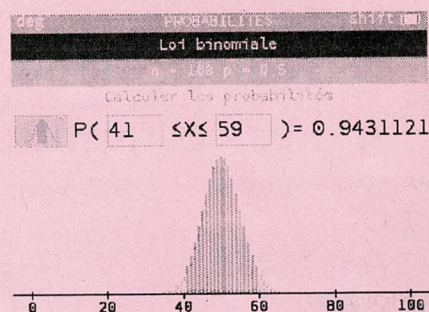
Exemple

On lance une pièce de monnaie équilibrée 100 fois de suites et X est la variable aléatoire qui compte le nombre de Pile obtenus.

- Lancer une pièce et regarder la face obtenue constitue une épreuve de Bernoulli dont le succès est « Obtenir Pile » de probabilité $p = 0,5$. On répète $n = 100$ fois cette épreuve de Bernoulli dans des conditions d'indépendance et on obtient un schéma de Bernoulli. X est la variable aléatoire qui compte le nombre de succès « Obtenir Pile ». Donc X suit la loi binomiale $B(100 ; 0,5)$.
- L'espérance de X est $\mu = np = 100 \times 0,5 = 50$ et sa variance est $V = np(1 - p) = 100 \times 0,5 \times 0,5 = 25$.
- D'après l'inégalité de Bienaymé-Tchebychev, pour tout réel $\delta > 0$: $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$.
Soit $P(|X - 50| \geq \delta) \leq \frac{25}{\delta^2}$. En particulier, si $\delta = 10$, alors $P(|X - 50| \geq 10) \leq 0,25$.
- Or, $P(|X - 50| < 10) = 1 - P(|X - 50| \geq 10)$, donc $P(|X - 50| < 10) \geq 0,75$.
- Cela signifie que la probabilité que le nombre de Pile soit dans l'intervalle $]40 ; 60[$ est supérieure ou égale à 0,75.

Remarque

On peut visualiser cette distribution des probabilités autour de la moyenne.

**Propriété : application à $\delta = k\sigma$**

Soient X une variable aléatoire d'espérance μ , de variance V et d'écart type σ et $k \in \mathbb{N}^*$.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Démonstration

On applique l'inégalité de Bienaymé-Tchebychev avec $\delta = k\sigma > 0$.

Donc $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$ soit $P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2}$ d'où $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Exemple

Appliquons l'inégalité de Bienaymé-Tchebychev dans différents cas :

$k = 2$	$k = 4$
$P(X - \mu \geq 2\sigma) \leq 0,25$	$P(X - \mu \geq 4\sigma) \leq 0,0625$
La probabilité que l'écart de X à μ soit supérieure ou égale à 2σ est inférieure ou égale à 0,25.	La probabilité que l'écart de X à μ soit supérieure ou égale à 4σ est inférieure ou égale à 0,0625.
On peut en déduire que les écarts de X à μ de quelques σ deviennent improbables.	

Exemple Simuler un échantillon de taille n d'une variable aléatoire

Au basket, lorsqu'elle tente un lancer franc, Claire marque le panier dans 70 % des cas.

À l'entraînement sur une série de 100 lancers, on note X le nombre de paniers réussis par Claire.

- Déterminer la loi de probabilité de X , son espérance μ et son écart type σ .

Tenter un panier franc est une épreuve de Bernoulli dont le succès est "le panier est réussi" de probabilité $p = 0.7$.

On répète $n = 100$ fois cette épreuve de Bernoulli de manière identique et dans des conditions d'indépendance.

On obtient un schéma de Bernoulli

X est la variable aléatoire qui compte le nombre de succès

X suit la loi binomiale de paramètres $n=100$ et $p=0,7$

$$\mu = E(X)$$

$$\sigma = \sigma(X)$$

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

$$\mu = 100 \times 0,7$$

$$\sigma = \sqrt{70 \times 0,3}$$

$$\mu = 70 \text{ parien}$$

$$\sigma = \sqrt{21} \text{ parien}$$

$$(\sigma \approx 4,58)$$

2. À l'aide de l'inégalité de Bienaymé-Tchebychev, justifier que $P(|X - \mu| \geq 2\sigma) \leq 0,25$.

$$\forall \delta \in \mathbb{R}^+ \quad P(|X - \mu| \geq \delta) \leq \frac{\sigma^2}{\delta^2}$$

$$\text{Ici, } \delta = 2\sigma \text{ donc } P(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{4\sigma^2}$$

$$\text{Ainsi: } P(|X - \mu| \geq 2\sigma) \leq 0,25$$

3. On a écrit ci-contre un programme en langage Python afin de simuler un échantillon de taille n de la variable aléatoire X .

- a. Interpréter le résultat renvoyé par la fonction Distance.

La fonction Distance renvoie l'écart

$|x - 70|$ où x est la valeur prise par

la variable aléatoire X

- b. Interpréter le résultat renvoyé par la fonction Echantillon.

Elle renvoie la proportion de valeurs de

x prise par X tel que $|x - 70| \geq 2\sigma$

```
1 from random import *
2
3 def Distance():
4     e=70
5     x=0
6     for k in range(100):
7         a=random()
8         if a<=0.7:
9             x=x+1
10    d=abs(x-e)
11    return d
12
13 def Echantillon(n):
14     s=4.58
15     y=0
16     for j in range(n):
17         if Distance()>=2*s:
18             y=y+1
19     p=y/n
20     return p
```

4. Saisir ce programme et exécuter Echantillon(1000) plusieurs fois.
5. Comparer au résultat obtenu par l'inégalité de Bienaymé-Tchebychev. Que peut-on observer ?

L'égalité de Bienaymé-Tchebychev est bien validée mais n'est pas optimale

II- LOI DES GRANDS NOMBRES

Propriété : inégalité de concentration

Soient :

- n un entier naturel supérieur ou égal à 2.
- $X_1 ; X_2 ; \dots ; X_n$ n variables aléatoires indépendantes et identiques qui suivent une même loi de probabilité d'espérance μ et de variance V constituant un échantillon $(X_1 ; X_2 ; \dots ; X_n)$.
- $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ la variable aléatoire moyenne de cet échantillon.

Pour tout réel $\delta > 0$, $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

Démonstration

Comme M_n est la moyenne de l'échantillon $(X_1 ; X_2 ; \dots ; X_n)$ de n variables aléatoires suivant la même loi d'espérance μ et de variance V , alors on peut utiliser les résultats de la séquence 13 :

- L'espérance de M_n est : $E(M_n) = \mu$;
- La variance de M_n est : $V(M_n) = \frac{V}{n}$.

On applique alors l'inégalité de Bienaymé-Tchebychev :

Pour tout réel $\delta > 0$, $P(|M_n - \mu| \geq \delta) \leq \frac{V(M_n)}{\delta^2}$, soit $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

Exemple

Dans une société de démarchage par téléphone, on estime que 40 % des personnes appelées répondent effectivement. La société appelle n personnes. Pour tout entier k tel que $1 \leq k \leq n$, la variable aléatoire X_k donne 1 si la k -ième personne appelée répond et 0 sinon.

- Les variables aléatoires $X_1 ; X_2 ; \dots ; X_n$ sont indépendantes, identiques, de même loi :
 - d'espérance $\mu = 0 \times 0,6 + 1 \times 0,4 = 0,4$;
 - de variance $V = 0,6 \times (0 - 0,4)^2 + 0,4 \times (1 - 0,4)^2 = 0,24$.

a	0	1
$P(X_k = a)$	0,6	0,4

- D'après l'inégalité de concentration, pour tout réel $\delta > 0$, $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

Soit $P(|M_n - 0,4| \geq \delta) \leq \frac{0,24}{n\delta^2}$.

- Dans le cas où $n = 1\,000$ et $\delta = 0,1$, on obtient : $P(|M_n - 0,4| \geq 0,1) \leq 0,024$.
- On dit que l'on obtient pour M_n une précision de 0,1 avec un risque de 0,024.
- Cela signifie que lorsqu'on appelle 1 000 personnes, la probabilité que le nombre de personnes qui répondent soit en dehors de l'intervalle]300 ; 500[est inférieure ou égale à 0,024.

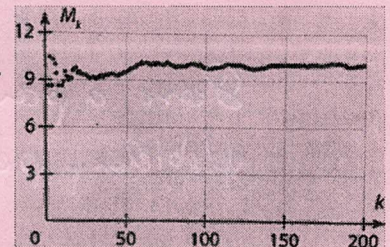
Propriété : loi des grands nombres

Soient :

- n un entier naturel supérieur ou égal à 2.
- $X_1 ; X_2 ; \dots ; X_n$ n variables aléatoires indépendantes et identiques qui suivent une même loi de probabilité d'espérance μ et de variance V constituant un échantillon $(X_1 ; X_2 ; \dots ; X_n)$.
- $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ la variable aléatoire moyenne de cet échantillon.

Pour tout réel $\delta > 0$, $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.DémonstrationPour tout réel $\delta > 0$ donné, l'inégalité de concentration permet d'écrire : $0 \leq P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.Or, $\lim_{n \rightarrow +\infty} \frac{V}{n\delta^2} = 0$, donc d'après le théorème des gendarmes, $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.Remarque

La loi des grands nombres illustre le fait que la moyenne de l'échantillon se rapproche de l'espérance des variables aléatoires quand la taille de l'échantillon « devient grande ».

Par exemple, dans l'échantillon $(X_1 ; X_2 ; \dots ; X_{200})$ de variables aléatoires d'espérance μ . M_k est la variable aléatoire moyenne de l'échantillon.On considère le nuage de points $(k ; M_k)$ ci-contre. M_k semble se stabiliser vers 10 lorsque k devient grand.On peut donc penser que $\mu = 10$.**Exemple Rechercher une taille d'échantillon**

Dans une ville moyenne de 20 000 habitants, lors d'une consultation portant sur la rénovation du théâtre municipal, 75 % des personnes consultées ont émis un avis positif.

On interroge n personnes et pour tout entier k tel que $1 \leq k \leq n$, la variable aléatoire X_k donne 1 si la k -ième personne interrogée est favorable au projet et 0 sinon.

1. Donner la loi de probabilité de X_k , son espérance μ et sa variance V .

Loi de probabilité de X_k :

Valeurs de X_k	0	1
Probabilité	0,25	0,75

Espérance de X_k :

$\mu = 0,75$

Variance de X_k :

$V = 0,1875$

2. On note M_n est la variable aléatoire moyenne de l'échantillon $(X_1; X_2; \dots; X_n)$.

Déterminer une taille n d'échantillon afin d'obtenir pour M_n une précision de 0,05 et un risque de 0,1.

D'après la loi de concentration:

$$\forall S \in \mathbb{R}^* \quad P(|M_n - \mu| \geq S) \leq \frac{V}{nS^2}$$

Ici, $\mu = 0,75; V = 0,1875; S = 0,05$

$$P(|M_n - 0,75| \geq 0,05) \leq \frac{0,1875}{n \times 0,05^2}$$

Pour $P(|M_n - 0,75| \geq 0,05) \geq \frac{75}{n} \leq 0,1$

On cherche $n \in \mathbb{N}$ tel que $P(|M_n - 0,75| \geq 0,05) \leq 0,1$

Il suffit de choisir n tel que $\frac{75}{n} \leq 0,1$: (I)

(I) $\Rightarrow n \geq 750$

Pour à partir d'un échantillon de taille 750 on est certain d'obtenir pour M_n une précision de 0,05 et un risque de 0,1.

3. Déterminer une taille n d'échantillon afin d'obtenir pour M_n une précision de 0,01 et un risque de 0,05.

Peut-on raisonnablement envisager cette situation ?

D'après 2)

Ici, $\mu = 0,75; V = 0,1875; S = 0,01$

$$P(|M_n - 0,75| \geq 0,01) \leq \frac{0,1875}{n \times 0,01^2}$$

Pour $P(|M_n - 0,75| \geq 0,01) \leq \frac{1875}{n}$

On cherche $n \in \mathbb{N}$ tel que $P(|M_n - 0,75| \geq 0,01) \leq 0,05$

Il suffit de choisir n tel que : $\frac{1875}{n} \leq 0,05$: (I')

(I') $\Rightarrow n \geq 37500$

Or, la ville compte 20K habitants

Pour un tel échantillon n'est pas possible