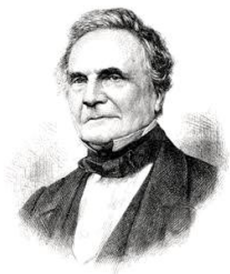
INEGALITE DE CONCENTRATION
ET
LOI DES GRANDS NOMBRES

Objectifs :

• *Appliquer l'inégalité de Bienaymé-Tchebychev pour définir une taille d'échantillon, en fonction de la précision et du risque choisi.*

1. Point historique

La parution de l'*Ars Conjectandi* de Jacques Bernoulli (1713), reprenant notamment d'anciens travaux de Huygens, marque une rupture dans l'histoire des probabilités. On y trouve la première étude de la distribution binomiale, introduite dans le cadre d'un tirage sans remise pour un modèle d'urne.
Un résultat majeur de cet ouvrage est son « théorème d'or », la loi des grands nombres, qui relie fréquences et probabilité, valide le principe de l'échantillonnage et est le premier exemple de « théorème limite » en théorie des probabilités. Le mathématicien français Bienaymé (en 1853, publication en 1867) et le mathématicien russe Tchebychev (en 1867) démontrent l'inégalité qui porte leur nom, en parlant de fréquences d'échantillons plutôt que de variables aléatoires. Ils fournissent ainsi la possibilité d'une démonstration plus simple de la loi des grands nombres.



Irénée-Jules Bienaymé
1796 – 1878

Pafnouti Tchebychev
1821 – 1894

1. Exploiter une moyenne d'échantillon

Définitions 1. ***Variable aléatoire moyenne de l'échantillon***

Soit $(X_1, X_2, \dots, X_n)$ un échantillon de taille n d'une loi de probabilité donnée.

On appelle variable aléatoire ……………… de l'échantillon la variable définie par :

$$M_n = \frac{X_1 + X_2 + \dots, +X_n}{n}.$$

Exemple : *Soit X la variable aléatoire qui, à chaque paquet de chips issue d'une chaîne de production, associe sa masse en grammes. On note X_i la variable aléatoire qui, à chaque lot de 3 paquets de chips, associe la masse du i-ème paquet.*

Les variables X_1, X_2 et X_3 sont indépendantes et suivent la même loi que X, donc (X_1, X_2, X_3) est un échantillon de taille .... de la loi X.

La variable aléatoire somme S_3 définie par $S_3 = X_1 + X_2 + X_3$ associe à .... sa masse en grammes.

La variable aléatoire moyenne $M_3 = \dfrac{X_1 + X_2 + X_3}{3}$ associe à chaque lot de 3 paquets la .... d'un paquet.

Propriété 1.

Soit n un entier naturel non nul et M_n la moyenne d'un échantillon de taille n de la loi suivie par une variable aléatoire X.

$$E(M_n) = ... \quad ; \quad V(M_n) = ... \quad \text{et } \sigma(M_n) = ... \quad .$$

Démonstrations :

$E(M_n) = E\left(\dfrac{X_1 + X_2 + ... + X_n}{n}\right) = ...$ (linéarité de l'espérance)

$= .... = ...$ (somme de v.a.)

$V(M_n) = V\left(\dfrac{X_1 + X_2 + ... + X_n}{n}\right) = ...$ (propriété de la variance)

$= .... = ...$ (somme de v.a.)

$\sigma(M_n) = ...$ (propriété l'écart-type)

$= .... = ...$

Exercice ❶

On considère la variable aléatoire X qui prend, de façon équiprobable, les valeurs -4 , 0, 1, 3 et 6.
M_{50} est la variable aléatoire moyenne d'un échantillon de taille 50 de la loi de X.
Calculer l'espérance, la variance et l'écart-type de M_{50}.

Exercice ❷

Dans un jeu de Scrabble, on compte 102 jetons parmi lesquels figurent deux jokers (valant 0 point) et les différentes lettres de l'alphabet dont les valeurs et les occurrences sont répertoriées dans le tableau suivant :

A_1	B_3	C_3	D_2	E_1	F_4	G_2	H_4	I_1	J_8	K_{10}	L_1	M_3
9	2	2	3	15	2	2	2	8	1	1	5	3

N_1	O_1	P_3	Q_8	R_1	S_1	T_1	U_1	V_4	W_{10}	X_{10}	Y_{10}	Z_{10}
6	6	2	1	6	6	6	6	2	1	1	1	1

1) Interpréter la variable aléatoire S_7 dans le cadre de l'exercice.
2) Déterminer $E(S_7)$ et interpréter le résultat obtenu.
3) On pose la variable aléatoire $M_7 = \frac{X_1 + X_2 + ... + X_7}{7}$.
Calculer $E(M_7)$ et interpréter le résultat obtenu.

source : site pigsels

Exercice ❸

La majorité des contraventions liées aux infractions au code de la route concernent les contraventions de 4e classe : utilisation du téléphone au volant, feu rouge grillé, non respect du port de la ceinture, chevauchement d'une ligne blanche continue, etc.
Le tarif forfaitaire d'une telle contravention s'élève à 135 €.
Il est cependant possible de payer la contravention moins chère si on procède rapidement au paiement.
Elle s'élève alors à 90 € et on parle d'amende minorée.
À l'inverse, en cas de retard de paiement, le prix de l'amende augmente et passe alors à 375 €. On parle alors d'amende majorée.
Une préfecture transmet les informations suivantes concernant le paiement des contraventions de 4e classe.

Montant de l'amende (en €)	90	135	375
Fréquence observée (en %)	79	15	6

Afin d'étudier de manière plus précise ces données, on prélève cent dossiers concernant les amendes de 4e classe. On suppose que le nombre d'amendes est suffisamment important pour assimiler cette expérience à un tirage avec remise.
Pour tout entier k compris de 1 à 100, (1 et 100 inclus), on appelle X_k la variable aléatoire correspondant au montant payé par le k^e dossier prélevé.

Exercice ❹

Un restaurant propose différents choix de menus.

• 64 % des clients ne souhaitent pas manger d'entrée, alors que 22 % des clients choisissent l'entrée à 9 €. Les autres choisissent l'entrée à 11 €.
• 76 % des clients commandent le plat composé de viande dont le prix s'élève à 19 €. Les autres choisissent le plat de poisson coûtant 22 €.
• Enfin, 10 % des clients mangent une crêpe en dessert à 8 €, 38 % des clients désirent manger le dessert au chocolat à 6 €, 30 % commandent le dessert aux fruits à 7 € et 22 % ne souhaitent pas manger de dessert.

On choisit un client du restaurant au hasard. On note respectivement X_1, X_2 et X_3 les variables aléatoires correspondant aux prix payés pour l'entrée, pour le plat et pour le dessert.

1) Déterminer les lois de probabilité des variables aléatoires X_1, X_2 et X_3.

2) On note X la variable aléatoire correspondant au prix total payé par le client.
Exprimer X en fonction de X_1, X_2 et X_3.

3) En déduire le prix moyen payé par chaque client.

4) On choisit maintenant dix clients de ce restaurant. On suppose que le nombre de clients du restaurant est suffisamment important pour assimiler cette expérience à un tirage avec remise.
En moyenne, sur ces dix clients, à combien le prix total s'élève-t-il ?

source : [site piqsels]

2. Inégalité de Bienaymé-Tchebychev

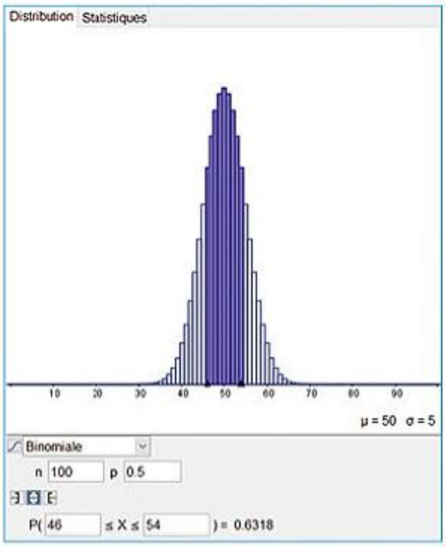
Activité : Soit X une variable aléatoire qui suit une loi binomiale de paramètres $n=100$ et $p=\frac{1}{2}$.

1) Déterminer l'espérance $E(X)$ et la variance $V(X)$ de X.

2) Dans l'inégalité de Bienaymé-Tchebychev, on compare pour tout réel strictement positif δ, $P(|X-E(X)|\geq\delta)$ et $\frac{V(X)}{\delta^2}$.

a) Compléter le tableau suivant à l'aide de la calculatrice ou d'un logiciel de calcul formel. *Arrondir les résultats au centième*.

δ	5	8	10	12
$P(\|X-E(X)\|\geq\delta)$	0,37			
$\frac{V(X)}{\delta^2}$	1			



b) Proposer une conjecture.

Propriété 2. Inégalité de Bienaymé-Tchebychev

Pour toute variable aléatoire X d'espérance $E(X)$ et de variance $V(X)$, pour tout réel δ strictement positif, $P(|X-E(X)|\geq\delta)......\frac{V(X)}{\delta^2}$.

Autrement dit : la probabilité que X se trouve en dehors de l'intervalle $]E(X)-\delta\,;\,E(X)+\delta[$ est inférieure à $\frac{V(X)}{\delta^2}$.

On dit que l'intervalle $\left]E(X)-\delta\ ;\ E(X)+\delta\right[$ est un intervalle de fluctuation de X.

Exemple : *Le taux moyen de glycémie dans une population est de 1 g.L^{-1} avec une variance de 0,1. Une personne présente un taux X critique si son taux ne se situe pas dans l'intervalle $]0,5\ ;1,5[$. Cet évènement se traduit par l'inégalité : $|X - E(X)| \geq$ ....*

Sa probabilité vérifie donc $P(|X - E(X)| \geq ....) \leq$ ...., c'est-à-dire

$P(|X - E(X)| \geq ....) \leq$ ....

On remarque que $P(|X-E(X)|\geq\delta)=1-$......................... . D'après l'inégalité de Bienaymé-Tchebychev, on a $P(|X-E(X)|\geq\delta)\leq\dfrac{V(X)}{\delta^2}$, alors $1-$......................... $\leq\dfrac{V(X)}{\delta^2}$

Donc $1-\dfrac{V(X)}{\delta^2}\leq$......................... . On en déduit alors :

Propriété 2 bis. Inégalité de Bienaymé-Tchebychev

Pour toute variable aléatoire X d'espérance $E(X)$ et de variance $V(X)$, pour tout réel δ strictement positif, $P(|X-E(X)|<\delta)$......$1-\dfrac{V(X)}{\delta^2}$.

Exercice ❺

Lors d'une saison de football, le nombre moyen de buts par match est de 2,5, avec une variance de 1,1. Majorer la probabilité que le match suivant ne se termine pas avec deux ou trois buts.

Exercice ❻

Dans une usine, la largeur moyenne d'une puce électronique prise au hasard est de 12mm et la variance est évaluée à 0,01.
Une puce n'est pas commercialisable si sa largeur est inférieure ou égale à 10,75 mm ou supérieure ou égale à 13,25 mm.
Majorer la probabilité qu'une puce ne soit pas commercialisable.

Exercice ❼

La taille moyenne d'une femme française est de 1,65 m et la variance des tailles des françaises est évaluée à 0,0025.
Jade affirme qu'il y a au plus un quart des femmes françaises dont la taille est inférieure ou égale à 1,55 m ou supérieure à 1,75 m.
Cette affirmation est-elle exacte ? (*On pourra considérer la variable aléatoire X prenant pour valeur la taille d'une femme choisie dans la population féminine française.*)

Exercice ❽

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

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

2) Recommencer avec $\delta = 3\sigma(X)$, puis $\delta = 4\sigma(X)$. Que constate-t-on ?

Exercice ❾

On interroge au hasard dix étudiants.
Les variables aléatoires $N_1, N_2, ..., N_{10}$ modélisent la note sur 20 obtenue à un examen par chacun d'entre eux. On admet que ces variables sont indépendantes et suivent la même loi binomiale de paramètres (20 ; 0,615).

1) Soit S la variable définie par $S = N_1 + N_2 + ... + N_{10}$.
Calculer l'espérance $E(S)$ et la variance $V(S)$ de la variable aléatoire S.

2) On considère la variable aléatoire $M = \dfrac{S}{10}$.

a) Que modélise cette variable aléatoire M dans le contexte de l'exercice ?
b) Justifier que $E(M) = 12{,}3$ et $V(M) = 0{,}47355$.
c) À l'aide de l'inégalité de Bienaymé-Tchebychev, justifier l'affirmation ci-dessous.
« La probabilité que la moyenne des notes de dix étudiants pris au hasard soit strictement comprise entre 10,3 et 14,3 est d'au moins 80 % ».

3. Inégalité de concentration

C'est Jacques Bernoulli qui publie l'une des premières versions de ce résultat dans son ouvrage posthume Ars Conjectandi en 1713. Il le démontre dans le cas particulier de la loi binomiale.

Jacques Bernoulli
1796 – 1878

Propriété 3. *Inégalité de concentration*

Soit la variable aléatoire moyenne M_n d'un échantillon de taille n de la variable aléatoire X. Pour tout réel δ strictement positif, on a : $P(|M_n - ....| \geq \delta) \leq \dfrac{.........}{.........}$

<u>*Démonstration*</u> : La variable aléatoire M_n a pour espérance $E(M_n) = ...$ et pour variance $V(M_n) = ...$. On applique l'inégalité de Bienaymé-Tchebychev à M_n : on

obtient : $P\left(\left|M_n - E\left(M_n\right)\right| \geq \delta\right) \leq \dfrac{V\left(M_n\right)}{\delta^2}$, ce qui équivaut à $P\left(\left|M_n - \text{.........}\right| \geq \delta\right) \leq \dfrac{\text{.........}}{\delta^2}$,

c'est-à-dire à $P\left(\left|M_n - \text{.........}\right| \geq \delta\right) \leq \dfrac{\text{.........}}{\text{.........}}$.

Exemple : On effectue n lancers successifs supposés indépendants d'une pièce équilibrée. On associe à chaque tirage i la variable aléatoire X_i prenant comme valeur 0 si on obtient face et 1 si on obtient pile. On pose $S_n = X_1 + X_2 + ... + X_n$ la variable aléatoire donnant le nombre de piles obtenu. On pose : $M_n = \dfrac{S_n}{n}$.

On a pour tout $i \in \{1\,;\,2\,;\,...\,;\,n\}$, $E\left(X_i\right) = \text{......}$ et $V\left(X_i\right) = \text{......}$.

Pour $n = 10\ 000$, l'inégalité de concentration donne : $P\left(\left|M_n - \text{......}\right| \geq 0{,}01\right) \leq \dfrac{\text{......}}{\text{.......................}}$

D'où $P\left(\left|M_n - \text{......}\right| \geq 0{,}01\right) \leq \text{......}$.

Ainsi, pour 10 000 lancers, la probabilité que la proportion de pile obtenue s'écarte de plus d'un centième de …… est inférieure à …….

Exercice ⑩

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 supérieure à 0,95.

Exercice 11

On lance une pièce équilibrée n fois de suite avec n un entier strictement positif.
La variable aléatoire M_n donne la proportion de pile obtenus au cours des n lancers.

1) Déterminer l'espérance et la variance de M_n.

2) Justifier que $P\left(\left|M_n - 0{,}5\right| \geq 0{,}1\right) \leq \dfrac{25}{n}$.

3) Déterminer une valeur de n telle que $P\left(\left|M_n - 0{,}5\right| \geq 0{,}1\right) \leq 0{,}05$.

On dit que M_n prend la valeur 0,5 avec une précision de 0,1 et un risque de 0,05.

Exercice 12

Le nombre de croissants vendus dans une boulangerie en une journée est une variable aléatoire d'espérance 50 et de variance 25.
Justifier que la probabilité que la vente de demain dépasse 75 croissants est inférieure à 0,04.

Exercice 13

100 personnes jouent indépendamment à un même jeu dont la variable aléatoire associée au gain (en euros) a pour espérance 10 et pour variance 2.
Donner une minoration de la probabilité que la moyenne des gains de ces 100 personnes soit comprise strictement entre 7 euros et 13 euros.

4. Loi faible des grands nombres

Propriété 4. *Loi faible des grands nombres*

Soit la variable aléatoire moyenne M_n d'un échantillon de taille n de la variable aléatoire X. Pour tout réel δ strictement positif, on a : $\lim\limits_{n \to +\infty} P\left(\left|M_n - E(X)\right| \geq \delta\right) = ...$

Remarques :
- Ce résultat justifie la possibilité de définir des probabilités en prenant pour valeurs approchées les fréquences obtenues pour un « grand » nombre d'essais.
- 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 : on lance n fois un dé équilibré à 8 faces et on nomme X_i la variable aléatoire donnant le résultat du i-ème lancer.
On admet que $E(X_i) = 4{,}5$ et $V(X_i) = 5{,}25$ pour tout entier i compris entre 1 et n.
Les lancers étant indépendants, $(X_1, X_2, ..., X_n)$ est un échantillon de variables aléatoires d'espérance 4,5 , de variance 5,25 et de moyenne $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$.
On considère $\delta = 0{,}1$.
D'après la loi des grands nombres, $P(|M_n - ......| \geq ......)$, que l'on peut également écrire $P(M_n \notin]......\,;\,......[)$, tend vers …… lorsque la taille de l'échantillon tend vers $+\infty$.
On en déduit que $P(M_n \in]......\,;\,......[)$ tend vers …… lorsque la taille de l'échantillon tend vers $+\infty$.
Autrement dit, si l'on fait un nombre suffisamment grand de lancers, on peut rendre l'événement « la moyenne de l'échantillon est dans $]......\,;\,......[$ » aussi probable qu'on le souhaite en prenant n suffisamment grand.

Exercice 14
Soit une variable aléatoire X d'espérance 3 et de variance 0,2.
Soit M_n la variable aléatoire moyenne d'un échantillon de taille n de la variable aléatoire X.
1) Déterminer un majorant de $P(|M_n - 3| \geq 0{,}1)$ pour $n = 100$, pour $n = 1\,000$, puis pour $n = 10\,000$. Que constate-t-on ?
2) Démontrer et interpréter le résultat précédent.

Exercice 15
On veut estimer l'espérance d'une variable aléatoire qui renvoie le rang du premier 6 lors d'une série de lancers d'un dé à 6 faces équilibré.
1) Compléter les fonctions Python ci-dessous en respectant la spécification donnée en documentation.

C. Lainé

```python
from random import randint

def premier6():
    """Renvoie le rang du premier 6
    lors d'une série de lancers de dés à 6 faces équilibrés"""
    n = 1
    while randint(1, 6) .....:
       n = .....
    return n

def echantillon_premier6(n):
    """Renvoie un échantillon de n réalisations de la va premier6"""
    return [............... for k in range(n)]

def moyenne(liste):
    """Renvoie la moyenne d'une liste de nombre"""
    s = 0
    for .................:
       s = .............
    return .............

def moyenne_echantillon_premier6(n):
    """Renvoie la moyenne sur un échantillon
    de taille n de la va premier6"""
return ......................
```

2) On a représenté sur le graphique ci-dessous des valeurs de `moyenne_echantillon_premier6(n)` pour des valeurs de n croissantes.
Quelle conjecture peut-on faire sur l'espérance de la variable aléatoire renvoyant le rang du premier 6 ?