

T spé

Inégalités de Markov et de Bienaymé-Tchebychev

Le 26-6-2023

Bernoulli
Ars conjectandi
galets blancs
Projet de document d'accompagnement Probabilités 17 mars 2008

Le 2-7-2023

- Simulation de variable aléatoire
- Interprétation de l'espérance d'une variable aléatoire X

Le 17-7-2023

Phrase écrite par Emma Picone vers la mi-juin en cours particulier

L'espérance d'une variable aléatoire X s'interprète comme la valeur moyenne prise par X lorsque l'on répète un grand nombre de fois l'expérience.

Le 27-4-2023

Paradoxe des anniversaires

Objectifs : Compléter un script. Utiliser des listes. Étudier un paradoxe.

1 On réunit un groupe de 30 personnes.
D'après vous, la probabilité p qu'au moins deux personnes du groupe soient nées le même jour est-elle inférieure à 0,2 ? Comprise entre 0,2 et 0,5 ? Supérieure à 0,5 ?

• Réponse :
https://manuelnumeriquemax.belin.education/cahier_python_pour_les_maths-seconde/topics/algo-2-t3c07-087-a_exercices

programme Python avec fonction double

Le 3-9-2022

Travail d'été passage en deuxième année ECG Lycée Montaigne

3. On lance indéfiniment une pièce de monnaie avec laquelle la probabilité d'obtenir *pile* est $a \in]0, 1[$ et on note Y la v.a.r. égal au nombre de lancers nécessaires pour obtenir la séquence *pile - pile*. On note, pour $i \in \mathbb{N}^*$: P_i : "le i -ème lancer donne un *pile*" et F_i : "le i -ème lancer donne un *face*".

(a) Calculer $P(Y = 2)$, $P(Y = 3)$ et $P(Y = 4)$.

(b) Soit k un entier ≥ 4 . Vérifier que $\{F_1, P_1 \cap P_2, P_1 \cap F_2\}$ est un SCE et en déduire que :

$$P(Y = k) = (1 - a)P(Y = k - 1) + a(1 - a)P(Y = k - 2) \quad (*)$$

(c) Compléter le programme Python suivant pour que l'appel de `proba(n, a)` (avec un entier $n \geq 4$ et un réel $a \in]0, 1[$) renvoie la valeur de $P(Y = n)$.

```
def proba(n,a)
    u = .....; v = .....
    for k in range (.....) :
        w = .....
        u = ..... ; .....
    return(w)
```

(d) On admet que Y possède une espérance. À l'aide de la relation (*), montrer que $E(Y) = \frac{1+a}{a^2}$.

Le mardi 28 juin 2022 (tapé le mercredi 13 juillet 2022)

Échantillonnage

① Poker (« test du poker »)

② Fluctuation d'échantillonnage

En fait, j'ai deux chapitres à faire après inégalités de Bienaymé-Tchebychev.

L'un avec le test du poker montre le principe des simulations à l'aide de nombres aléatoires.

L'autre montre des réalisations de simulations évidemment à actualiser avec Python.

Le 9-7-2022

« Des outils pour simuler » contient le mot Fankston

Le 7-7-2022

Lois des grands nombres
Prof. Mohamed El Merouani
Université Abdelmalek Essaâdi
Faculté des Sciences de Tétouan
Département de Mathématiques
<https://elmerouani.jimdofree.com/calcul-des-probabilités>
2020/2021

Téléchargé le 9-7-2022

Loi faible des grands nombres :

Remarque 1 :

En particulier, les hypothèses du théorème précédent sont toutes satisfaites si les v.a. X_n sont indépendantes et de même loi et si X_1 admet un moment d'ordre deux.

On étudie un cas particulier du théorème précédent (il lui est toutefois historiquement antérieur), c'est le théorème de Bernoulli.

Théorème de Bernoulli :

Soit une suite de n épreuves de Bernoulli indépendantes et soit X la v.a. égale au nombre de succès au cours de ces n épreuves. Alors $X \sim \mathcal{B}(n, p)$, où p est la probabilité de succès.

Prof. Mohamed El Merouani (Université Abdelmalek Essaâdi Faculté des Sciences de Tétouan Département de Mathématiques <https://elmerouani.jimdofree.com/calcul-des-probabilités>) Probabilités 2020/2021 8 / 27 Lois des grands nombres : Loi faible des grands nombres

Loi faible des grands nombres :

Preuve du théorème de Bernoulli :

On sait que X suit la loi binomiale de paramètres n et p : $X \sim \mathcal{B}(n, p)$.

Donc $E(X) = np$ et $\text{Var}(X) = npq$ avec $q = 1 - p$.

Soit la v.a. $Y = \frac{X}{n}$, alors $E(Y) = p$

$E(X) = np$

n

$= p$ et

$\text{Var}(Y) = E\left(\frac{X-np}{n}\right)^2 = \frac{1}{n^2} E((X-np)^2) = \frac{\text{Var}(X)}{n^2} = \frac{npq}{n^2}$

En utilisant l'inégalité de Bienaymé-Tchebychev : $P(|Y - p| \geq \varepsilon) \leq \frac{pq}{n\varepsilon^2}$

Comme nous ne connaissons pas la valeur de p , il sera nécessaire de calculer une valeur extrême supérieure de

$\frac{pq}{n\varepsilon^2}$.

d

dp

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2} = 0 \Leftrightarrow p =$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2} < 0$.

Donc $p = \frac{1}{2}$ est un maximum et par conséquent $\frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2} =$

$\frac{pq}{n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

d'où $P(|Y - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$ ou encore $P(|Y - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}$

$\frac{pq}{n\varepsilon^2}$

Par conséquent $\lim_{n \rightarrow +\infty} P(|Y - p| < \varepsilon) = 1$ et $X/n \xrightarrow{P} p$. \square

Prof. Mohamed El Merouani (Université Abdelmalek Essaâdi Faculté des Sciences de Tétouan Département de Mathématiques <https://elmerouani.jimdofree.com/calcul-des-probabilités>) Probabilités 2020/2021 9 / 27 Lois des grands nombres : Loi faible des grands nombres

Théorème de Bernoulli :

Ce théorème veut dire que nous pouvons utiliser X/n pour estimer la probabilité p , puisque c'est une estimation exacte lorsque n est suffisamment grande.

Dans la pratique, le nombre n doit être fini, par conséquent, on doit déterminer le nombre n de façon que, avec une probabilité suffisamment grande, la fréquence observée X/n soit comprise entre $p - \varepsilon$ et $p + \varepsilon$.

La fréquence observée des piles dans un jeu de pile ou face, après n tirages, est "proche" de la probabilité (déterministe) p d'obtenir pile, pourvu que n soit grande.

Donc, si p est inconnue (par exemple, on ne sait pas si la pièce est truquée), nous avons là un moyen de l'approximer.

Le 9-7-2022

Exercice 9
On considère un jeu dont le gain algébrique en euros est donné par la variable aléatoire G de loi suivante :

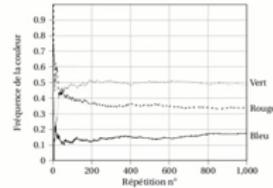
g_i	-2	-1	0	10
$P(G = g_i)$	0,33	0,44	0,22	0,01

De quelle valeur va se rapprocher la moyenne des gains d'un nombre important de joueurs jouant de manière indépendante ? Justifier.

$E(G) = 0,33 \times (-2) + 0,44 \times (-1) + 0,22 \times 0 + 0,01 \times 10 = -1$

Exercice 10

Dans un urne, il y a des boules rouges, des boules bleues et des boules vertes indiscernables au toucher. On réalise 1000 fois l'expérience aléatoire consistant à tirer au sort une boule dans l'urne, noter sa couleur et la remettre dans l'urne. L'évolution des fréquences des boules obtenues de chaque couleur au fil de ces 1000 répétitions est donnée ci-dessous :



On note p la probabilité d'obtenir une boule rouge et on note X la variable aléatoire qui, au résultat d'un tirage, associe 1 si on a pioché une boule rouge et 0 sinon. On note M_n la variable aléatoire moyenne d'un échantillon de taille n de la loi de X .

1. Donner $E(X)$ puis écrire la loi des grands

nombre en fonction de M_n et p .

2. À l'aide du graphique, estimer la valeur de p en justifiant.
3. Sachant qu'il y a 50 boules au total dans l'urne, estimer le nombre de boules rouges puis le nombre de boules vertes et bleues.

Exercice 11

On considère une urne contenant dix boules dont exactement 4 sont rouges. On tire successivement et avec remise une boule de l'urne et on s'arrête dès que l'on obtient une première boule rouge. L'objectif est de conjecturer l'espérance de cette loi à l'aide du programme informatique suivant :

```
import random

def rang(n):
    S = 0
    for i in range(n):
        X = random.randint(1,10)
        C = 1
        while X <= .... :
            X = random.randint(1,10)
            C = ....
        S = S + C
    return S
```

1. Reproduire et compléter le programme ci-dessus afin qu'il simule cette expérience.
2. Quel est le rôle de la variable S ?
3. Tester le programme pour de grandes valeurs de n . Quelle semble être l'espérance de cette loi? Justifier en citant le théorème utilisé.

Concentration, loi des grands nombres

Une urne contient 3 boules rouges et 2 boules vertes. On tire au hasard successivement et avec remise deux boules de l'urne.

1. X est la variable aléatoire qui donne le nombre de boules rouges tirées.
 - (a) Donner la loi de probabilité de X .
 - (b) Déterminer l'espérance et la variance de X .
2. On répète n fois l'expérience. X_k donne le nombre de boules rouges tirées lors de la k -ième expérience ($1 \leq k \leq n$). On pose $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$.
 - (a) Que représente la variable M_n . Déterminer son espérance et sa variance.
 - (b) Justifier que pour tout nombre $\delta > 0$, $P(|M_n - 1,2| \geq \delta) \leq \frac{0,48}{n\delta^2}$.
 - (c) Écrire l'inégalité précédente dans le cas où $\delta = 0,1$ et $n = 100$. Interpréter.
 - (d) On prend $\delta = 0,01$. Choisir alors une valeur de n telle que $P(|M_n - 1,2| \geq 0,01) \leq 0,48$. Donner une interprétation de cette valeur.

Le 2 juin 2022

Du point de vue historique, le résultat fondamental permettant de comprendre la façon dont se stabilisent les fréquences des résultats au fur et à mesure que le nombre des essais augmente est le théorème de Bernoulli²⁰. Ce théorème, qui confirme l'intuition sur la stabilité des fréquences, est plus difficile à démontrer que la formule de Bayes. La démonstration « moderne » dans la théorie axiomatique de Kolmogorov repose sur une majoration assez grossière donnée par la formule de Bienaymé-Tchebychev, majoration peu utile en pratique.

Document probabilités accompagnement mars 2016

Vidéos

9:46

Exercices : Loi des grands nombres

YouTube · Mathemax

9 minutes et 46 secondes

7 juin 2021

Inégalités de Markov et de Bienaymé-Tchebychev

Plan du chapitre :

I. Inégalité de Markov

II. Inégalité de Bienaymé-Tchebychev

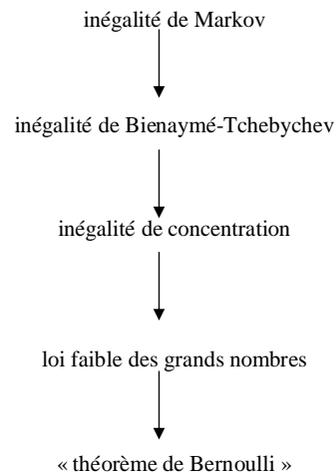
III. Inégalité de concentration

IV. La loi faible des grands nombres

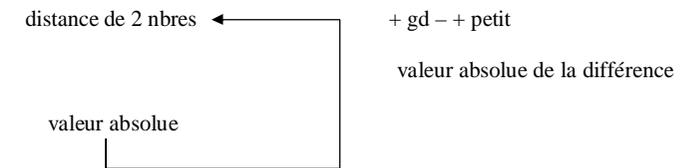
V. Échantillonnage et estimation

VI. Application au cas d'une loi binomiale

Progression du chapitre



Dans tout le chapitre, on se place dans un espace probabilisé fini (Ω, P) .



Valeur absolue et intervalle

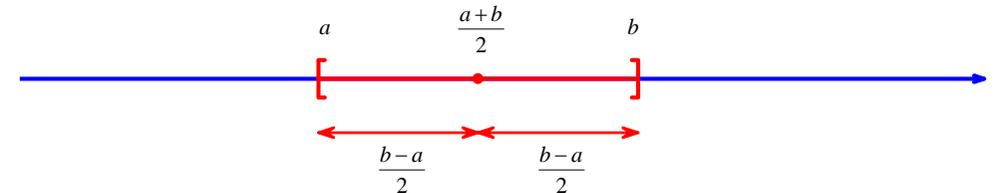
1°) Centre et rayon d'un intervalle $[a ; b]$ ($a \leq b$)

• Définitions

★ Centre de l'intervalle $[a ; b]$: $c = \frac{a+b}{2}$

★ Longueur de l'intervalle $[a ; b]$: $b - a$ (différence entre la borne supérieure et la borne inférieure)

★ Rayon de l'intervalle $[a ; b]$: $r = \frac{b-a}{2}$



• Exemple : $[1 ; 7]$

Centre : 4
Longueur : 6
Rayon : 3

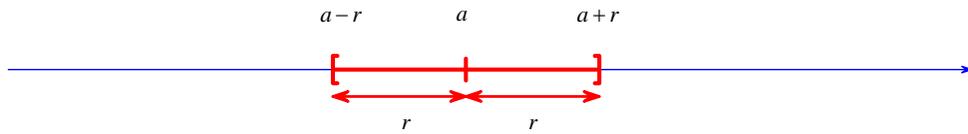
• Caractérisation de l'appartenance d'un élément à l'aide de la valeur absolue (cf. 2°) :

$$x \in [a ; b] \text{ équivaut à } \left| x - \frac{a+b}{2} \right| \leq \frac{b-a}{2}$$

• Remarque de vocabulaire :

On parle du centre de l'intervalle $[a ; b]$ et non de son milieu.

2°) Lien entre intervalles et valeur absolue



x appartient à l'intervalle de centre a et de rayon r ($r > 0$) équivaut à $|x - a| \leq r$.

On retient :

$$|x - a| \leq r \text{ équivaut à } a - r \leq x \leq a + r$$

Exemple :

$$|x - 2| \leq 3 \text{ équivaut à } 2 - 3 \leq x \leq 2 + 3 \\ \text{équivaut à } -1 \leq x \leq 5$$

N.B. : La notion de centre d'un intervalle a déjà été abordée en statistiques avec la notion de **centre de classe**.

intervalle borné non vide

ouvert ou fermé

centre : moyenne des extrémités

rayon : différence des extrémités divisée par 2

$\mathbb{R} \rightarrow$ pas de centre

• La distance de deux réels x et y est donnée par la formule $d(x; y) = |x - y| = |y - x|$.

• (Exemple : $d(u_n; l) = |u_n - l|$)

• $|x - a| < \varepsilon \Leftrightarrow a - \varepsilon < x < a + \varepsilon$

(Exemple : $|u_n - l| < \varepsilon \Leftrightarrow l - \varepsilon < u_n < l + \varepsilon$)

On peut reformuler la définition avec une valeur absolue :

(u_n) tend vers l signifie que pour tout réel $\varepsilon > 0$, il existe un entier naturel N tel que si $n \geq N$, alors

$$|u_n - l| < \varepsilon.$$

I. Inégalité de Markov

1°) Propriété [inégalité de Markov]

Soit X une variable aléatoire réelle sur Ω à valeurs positives ou nulles.

Alors pour tout réel $a > 0$, on a $P(X \geq a) \leq \frac{E(X)}{a}$.

Cette inégalité a été nommée ainsi en l'honneur d'Andreï Markov.

Andreï Andreïevitch Markov (1856-1922) est un mathématicien russe.

Il est considéré comme le fondateur de la théorie des processus stochastiques.

Commentaires :

- La condition « X est à valeurs positives ou nulles » est essentielle.
- La loi de probabilité de X n'importe pas pour pouvoir appliquer cette inégalité.

On sait que si X est une variable aléatoire à valeurs positives ou nulles, alors son espérance est aussi positive ou nulle.

- Si on applique l'inégalité de Markov pour un réel $a \leq E(X)$, alors $\frac{E(X)}{a} \geq 1$.

On obtient donc une majoration de $P(X \geq a)$ par un réel supérieur ou égal à 1, ce qui ne présente pas d'intérêt puisque l'on sait qu'une probabilité est inférieure ou égale à 1.

L'inégalité de Markov ne donne une majoration intéressante de $P(X \geq a)$ que lorsqu'on l'applique avec un réel $a > E(X)$.

- L'inégalité de Markov permet de démontrer que $P(X \geq a) \xrightarrow{a \rightarrow +\infty} 0$ (logique).

On applique pour cela le théorème des gendarmes car $0 \leq P(X \geq a) \leq \frac{E(X)}{a}$ pour tout réel $a > 0$ et

$$\frac{E(X)}{a} \xrightarrow{a \rightarrow +\infty} 0.$$

En fait, on a supposé que X est une variable aléatoire qui prend un nombre fini de valeurs (car l'univers de l'expérience aléatoire est fini) donc si a est supérieur à la plus grande valeur possible prise par X , alors

$P(X \geq a) = 0$ (probabilité d'un événement impossible).

2°) Exemple

Une usine produit en moyenne 35 pièces par semaine. On note X la variable aléatoire donnant le nombre de pièces produites par semaine.

Que peut-on dire de la probabilité que l'usine produise plus de 70 pièces par semaine ?

D'après le contexte, on sait que X est à valeurs positives ou nulles et que $E(X) = 35$ (on assimile l'espérance à la moyenne).

D'après l'inégalité de Markov, $P(X \geq 70) \leq \frac{35}{70}$ soit $P(X \geq 70) \leq \frac{1}{2}$.

$\frac{1}{2}$ est un majorant de $P(X \geq 70)$.

3°) Démonstration

On sait que $E(X) = \sum_{\omega \in \Omega} X(\omega) \times P(\{\omega\})$ (formule donnant l'espérance d'une variable aléatoire).

On note A l'événement constitué des éléments $\omega \in \Omega$ tels que $X(\omega) \geq a$.

Autrement dit, $A = \{X \geq a\}$.

Comme tous les termes de la somme sont positifs ou nuls, $E(X) \geq \sum_{\omega \in A} X(\omega) \times P(\{\omega\})$.

On a $\sum_{\omega \in A} (X(\omega) \times P(\{\omega\})) \geq \sum_{\omega \in A} (a \times P(\{\omega\}))$.

Or $\sum_{\omega \in A} (a \times P(\{\omega\})) = a \times \sum_{\omega \in A} P(\{\omega\}) = a \times P(A)$.

On a donc $E(X) \geq a \times P(A)$.

Comme a est strictement positif par hypothèse, cette dernière inégalité est équivalente à $P(A) \leq \frac{E(X)}{a}$ qui est l'inégalité de Markov.

II. Inégalité de Bienaymé-Tchebychev

On s'intéresse à l'écart entre une variable aléatoire X et son espérance.

1°) Théorème [inégalité de Bienaymé-Tchebychev]

Soit X une variable aléatoire réelle sur Ω .

Pour tout réel $a > 0$, on a : $P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$.

Irénée-Jules Bienaymé (1796-1878) est un mathématicien français qui a énoncé cette inégalité en 1853. C'est le mathématicien russe Pafnouti Tchebychev (1821-1894) qui l'a démontrée en 1867.

Bienaymé et Tchebychev (revue Tangente décembre 2020)

Pafnouti Lvovitch Tchebychev (1821-1894), le plus célèbre mathématicien russe du XIX^e siècle, ne se prénommait pas Bienaymé comme le pensent parfois certains étudiants. Si les noms de ces deux mathématiciens sont associés pour désigner une célèbre inégalité, ce n'est pas un hasard. Irénée-Jules Bienaymé (1796-1878) était un respectable inspecteur des finances lorsqu'il fut exclu de son poste en 1848 pour « manque de chaleur républicaine ». Il se tourne alors vers une brillante carrière de mathématicien qui le conduira à l'Académie des sciences. Il fait la connaissance de Tchebychev en octobre 1852 et une forte amitié s'installe entre les deux hommes. Le mathématicien russe séjourne plusieurs fois chez le savant français et celui-ci, féru de langue russe, traduit les écrits de son ami, ce qui permet à celui-ci de diffuser ses recherches. 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.

Autres notations :

On note m l'espérance de X et σ l'écart-type de X.

On a donc $E(X) = m$ et $V(X) = \sigma^2$.

L'inégalité de Bienaymé-Tchebychev s'écrit $P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}$.

Si on applique l'inégalité de Bienaymé-Tchebychev pour un réel $a \leq \sigma$, alors $\frac{\sigma}{a} \geq 1$ et par conséquent, $\frac{\sigma^2}{a^2} \geq 1$ (par élévation au carré des deux membres, sachant que tous les deux sont positifs).

On obtient donc une majoration de $P(|X - m| \geq a)$ par un réel supérieur ou égal à 1, ce qui ne présente pas d'intérêt puisque l'on sait qu'une probabilité est inférieure ou égale à 1.

L'inégalité de Bienaymé-Tchebychev ne donne une majoration intéressante de $P(|X - m| \geq a)$ que lorsqu'on l'applique avec un réel $a > \sigma$.

2°) Démonstration

On a l'équivalence fondamentale suivante : $|X - m| \geq a \Leftrightarrow |X - m|^2 \geq a^2$ (car les deux membres sont positifs ou nuls).

Or $Y = (X - m)^2$ est une variable aléatoire à valeurs positives ou nulles.

Donc par inégalité de Markov, $P(Y \geq a^2) \leq \frac{E(Y)}{a^2}$.

Or $E(Y) = V(X)$ d'où l'inégalité.

3°) Exercices

① Le taux moyen de glycémie est de 1 g.L^{-1} avec une variance de $0,1 \text{ (g.L}^{-1})^2$.

Une personne présente un taux critique si son taux ne se situe pas dans l'intervalle $]0,5 ; 1,5[$.

Déterminer un majorant de la probabilité qu'une personne présente un taux critique.

On note X la variable aléatoire qui modélise la glycémie en g.L^{-1} .

On a $E(X) = 1$ et $V(X) = 0,1$.

L'intervalle $]0,5 ; 1,5[$ a pour centre 1 et pour rayon 0,5. On peut écrire $]0,5 ; 1,5[=]1 - 0,5 ; 1 + 0,5[$.

$X \in]0,5 ; 1,5[\Leftrightarrow |X - 1| < 0,5$

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X - 1| \geq 0,5) \leq \frac{0,1}{0,5^2}$ soit $P(|X - 1| \geq 0,5) \leq 0,4$.

La probabilité qu'une personne présente un taux critique est inférieure ou égale à 0,4.

② On lance 100 fois une pièce équilibrée.

Déterminer à l'aide de l'inégalité de Bienaymé-Tchebychev un minorant de la probabilité que le nombre de piles soit compris dans l'intervalle $I =]40 ; 60[$.

On note X la variable aléatoire qui compte le nombre de piles obtenus à l'issue des 100 lancers.

On cherche à majorer $P(X \in I)$.

X suit la loi binomiale de paramètres $n = 100$ (nombre de répétitions) et $p = \frac{1}{2}$ (probabilité d'obtenir pile en un lancer).

On calcule l'espérance et la variance de X .

$$E(X) = 100 \times \frac{1}{2} = 50 \quad (\text{formule de l'espérance d'une variable aléatoire qui suit une loi binomiale})$$

$$V(X) = 100 \times \frac{1}{2} \times \frac{1}{2} = 25 \quad (\text{formule de la variance d'une variable aléatoire qui suit une loi binomiale})$$

L'intervalle I a pour centre 50 et pour rayon 10.

Le fait que X appartienne à I se traduit par $|X - 50| < 10$.

Nous allons transformer en utilisant l'événement contraire.
Pour cela, on écrit : $P(|X - 50| < 10) = 1 - P(|X - 50| \geq 10)$.

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X - 50| \geq 10) \leq \frac{25}{10^2}$ soit $P(|X - 50| \geq 10) \leq \frac{1}{4}$.

On multiplie les deux membres de cette dernière inégalité par -1 .

On obtient $-P(|X - 50| \geq 10) \geq -\frac{1}{4}$.

En ajoutant ensuite 1 à chaque membre de l'inégalité, on obtient $1 - P(|X - 50| \geq 10) \geq 1 - \frac{1}{4}$ ce qui donne

finalement $P(|X - 50| < 10) \geq \frac{3}{4}$ soit $P(X \in I) \geq \frac{3}{4}$.

La probabilité que le nombre de piles soit dans l'intervalle I est supérieure ou égale à 0,75.

Complément :

La probabilité que X soit compris dans I est égale à 0,943... (utilisation de la commande de la calculatrice permettant de calculer des probabilités avec une loi binomiale).

$$P(X \in I) = P(40 < X < 60) = P(41 \leq X \leq 59) = 0,9431120663590...$$

Pour effectuer le calcul sur calculatrice, on peut éventuellement écrire $P(X \in I) = P(X \leq 59) - P(X \leq 40)$.

4°) Situation particulière de l'inégalité de Bienaymé-Tchebychev

On considère une variable aléatoire X d'espérance m et d'écart type σ .

Nous allons majorer grâce à l'inégalité de Bienaymé Tchebychev les probabilité que $|X - m| \geq \sigma$,
 $|X - m| \geq 2\sigma$, $|X - m| \geq 3\sigma$.

• L'inégalité de Bienaymé-Tchebychev donne $P(|X - m| \geq \sigma) \leq \frac{\sigma^2}{\sigma^2}$ soit $P(|X - m| \geq \sigma) \leq 1$.

Cette inégalité ne présente pas d'intérêt car elle toujours vraie (une probabilité est toujours inférieure ou égale à 1).

• L'inégalité de Bienaymé-Tchebychev donne $P(|X - m| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2}$ soit $P(|X - m| \geq 2\sigma) \leq \frac{1}{4}$.

Cette inégalité est déjà plus intéressante.

On peut en déduire que $P(|X - m| < 2\sigma) \geq \frac{3}{4}$ ou encore $P(X \in]m - 2\sigma ; m + 2\sigma[) \geq \frac{3}{4}$.

Par conséquent, on a la minoration suivante $P(X \in [m - 2\sigma ; m + 2\sigma]) \geq \frac{3}{4}$.

Cette minoration permet de mieux comprendre le rôle l'écart-type comme mesure de dispersion des valeurs autour de l'espérance.

L'intervalle $[m - 2\sigma ; m + 2\sigma]$ joue un rôle important en probabilités.

• L'inégalité de Bienaymé-Tchebychev donne $P(|X - m| \geq 3\sigma) \leq \frac{\sigma^2}{(3\sigma)^2}$ soit $P(|X - m| \geq 3\sigma) \leq \frac{1}{9}$.

Cette inégalité est déjà plus intéressante.

L'inégalité de Bienaymé-Tchebychev montre donc que des écarts de X à $E(X)$ de quelques σ deviennent improbables.

Le caractère universel de l'inégalité de Bienaymé-Tchebychev a pour contrepartie le fait qu'elle est loin d'être optimale. Par exemple, on a démontré que $P(|X - m| \geq 2\sigma) \leq \frac{1}{4}$ mais on constate par simulation que cette probabilité est souvent majorée par 0,05 (travail fait en première).

III. Inégalité de concentration

1°) Propriété [inégalité de concentration]

Soit (X_1, X_2, \dots, X_n) un échantillon de variables aléatoires d'espérance m et d'écart-type σ .

On pose $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ (moyenne de l'échantillon).

Alors, pour tout réel a strictement positif, $P(|M_n - m| \geq a) \leq \frac{\sigma^2}{na^2}$.

2°) Démonstration

On sait que $E(M_n) = m$ et $V(M_n) = \frac{\sigma^2}{n}$ (propriété vue dans le chapitre précédent).

On applique ensuite l'inégalité de Bienaymé-Tchebychev.

On obtient directement : $P(|M_n - m| \geq a) \leq \frac{\sigma^2}{na^2}$.

IV. La loi faible des grands nombres

1°) Théorème [loi faible des grands nombres]

Soit (X_1, X_2, \dots, X_n) un échantillon de variables aléatoires d'espérance m et d'écart-type σ .

On pose $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ (moyenne de l'échantillon).

Alors, pour tout réel a strictement positif, $P(|M_n - m| \geq a) \xrightarrow{n \rightarrow +\infty} 0$.

La loi faible des grands nombres illustre d'une certaine manière le fait que la moyenne de l'échantillon se rapproche de l'espérance des variables aléatoires quand la taille de l'échantillon « devient grande », comme cela a été vu en 1^{ère}.

Il existe une loi forte des grands nombres qui n'est pas au programme.

2°) Démonstration

On utilise l'inégalité de concentration déduite de l'inégalité de Bienaymé-Tchebychev.

Cette inégalité s'écrit : $P(|M_n - m| \geq a) \leq \frac{\sigma^2}{na^2}$.

On peut écrire $0 \leq P(|M_n - m| \geq a) \leq \frac{\sigma^2}{na^2}$.

Comme $\frac{\sigma^2}{na^2} \xrightarrow{n \rightarrow +\infty} 0$, par théorème des gendarmes, on en déduit que $P(|M_n - m| \geq a) \xrightarrow{n \rightarrow +\infty} 0$.

L'espérance d'une variable aléatoire X s'interprète comme la valeur moyenne prise par X lorsqu'on répète un très grand nombre de fois l'expérience.

V. Échantillonnage et estimation

Le samedi 3 septembre 2022

T spé 22 h 37

Pour chaque échantillon, la fréquence observée est plus ou moins proche de p .

On peut évaluer la distance entre f et p en calculant $|f - p|$.

Attention au sens du mot échantillon (différente de celui d'échantillon d'une variable aléatoire).

1°) Intervalle de fluctuation d'une fréquence

Cette notion, historiquement, apparaît dans les premiers travaux de Jacques Bernoulli (1654-1703).

Après sa mort, on trouva dans ses papiers des calculs extrêmement poussés qu'il n'avait osé publier de son vivant.

Exemple :

On lance 100 fois une pièce de monnaie équilibrée. Il est très peu probable d'obtenir 50 piles et 50 faces. En revanche, il est intéressant de déterminer quelles sont les valeurs les plus probables du nombre de piles. Il est intéressant de déterminer une « fourchette » dans laquelle va se trouver le nombre de piles dans la plupart des cas.

On note X le nombre de piles obtenus à l'issue des 100 lancers.

On est dans le cas d'un schéma de Bernoulli.

La variable X suit donc la loi binomiale de paramètres $n = 100$ (nombre de répétitions) et $p = \frac{1}{2}$ (probabilité d'obtenir pile, puisque la pièce est équilibrée).

On peut démontrer que la probabilité que $X \in \llbracket 40 ; 60 \rrbracket$ (il s'agit de l'intervalle d'entiers allant de 40 à 60, c'est-à-dire de l'ensemble des entiers naturels de 40 à 60 inclus : 40, 41, 42, ..., 59, 60) avec une probabilité proche de 0,95 (la probabilité est égale à 0,964799...).

On dit que l'intervalle d'entiers $\llbracket 40 ; 60 \rrbracket$ est un intervalle de fluctuation du nombre de piles au seuil approximatif de 95 %.

Formule générale donnant un intervalle de fluctuation d'une fréquence

$$\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$$

Il s'agit d'un intervalle de fluctuation centré de centre p .

Les conditions d'application sont ...

Plus n est grand, plus l'intervalle est resserré.

Plus on lance, plus se concentre.

2°) Intervalle de fluctuation d'une moyenne

En 1^{ère}, on a étudié les problèmes d'échantillonnage de moyenne (« moyenne d'un échantillon ») en utilisant des simulations.

Propriété (admise sans démonstration)

On considère une population avec un caractère qualitatif C.
On note μ la moyenne et σ l'écart-type du caractère C dans la population.
On constitue N échantillons aléatoires de taille n .
Si m désigne la moyenne d'un échantillon, la proportion des cas où l'écart entre m et μ est inférieur ou égal à $\frac{2\sigma}{\sqrt{n}}$ est proche de 95 %.

Cela traduit que la proportion d'échantillons pour lesquels on a $|m - \mu| \leq \frac{2\sigma}{\sqrt{n}}$ (utilisation de la valeur absolue pour traduire l'écart) ou encore $m \in \left[\mu - \frac{2\sigma}{\sqrt{n}}; \mu + \frac{2\sigma}{\sqrt{n}} \right]$ est proche de 95 %.

On dit que l'intervalle $\left[\mu - \frac{2\sigma}{\sqrt{n}}; \mu + \frac{2\sigma}{\sqrt{n}} \right]$ est un intervalle de fluctuation de la moyenne au seuil approximatif de 95 % pour des échantillons aléatoires de taille n .

3°) Estimer une proportion

Dans de nombreuses situations, on est amené à estimer la proportion d'un caractère observé sur une population à partir d'un échantillon, **par exemple les sondages**. On dispose d'outils mathématiques permettant d'estimer la marge d'erreur de procédés.

Les estimations sont essentielles dans l'étude de populations, en sciences sociales ou en santé publique.

Lors de l'étude d'un caractère sur un échantillon prélever dans une population.

Un intervalle de confiance permet d'estimer entre quelles bornes la proportion réelle sur toute la population pourrait se trouver.

Constituer un échantillon pour estimer une proportion revient à étudier des sommes de variables aléatoires.

La taille n de cet échantillon amène à obtenir une suite de variables aléatoires sommes.

Plus la taille de l'échantillon est grande, plus l'estimation est précise : c'est la loi des grands nombres.

Certains échantillons « biaisés » ne reflètent pas la réalité de la population : des exemples existent, notamment lors de sondages électoraux.

Les instituts de sondages préfèrent la méthode des quotas afin de sélectionner des échantillons plus fidèles à la structure d'une population ; dans certains cas, cependant, on peut utiliser la méthode aléatoire.

Le lundi 30 mai 2022

En enseignement scientifique, en lien avec les populations, les élèves ont étudié :

intervalle de fluctuation

intervalle de confiance

fluctuare : mot latin lié aux bateaux (être battu par les flots) qui vient lui-même de fluctus (qui a donné les mots flots et flux en français)

fluctuation

Au premier trimestre, les notes d'un élève ont fluctué entre 0 et 5.

Fluctuations du cours de la bourse

Opinion fluctuante

Humeur fluctuante, états d'âme fluctuants

Il existe des méthodes pour déterminer des intervalles de fluctuation (par calcul ou autre).

Estimation d'une espérance

Si on fait 10 expériences et que l'on note T_1, T_2, \dots, T_{10} les nombres de réels qu'il a fallu choisir pour avoir une somme qui dépasse 1, $\frac{T_1 + T_2 + \dots + T_{10}}{10}$ donne une estimation du nombre moyen de nombres choisis.

Le 4-6-2022

Estimation d'une espérance

Loi des grands nombres

10. Utiliser un (bon) simulateur aléatoire pour simuler N réalisations $(x_i)_{0 \leq i < N}$ d'une variable aléatoire X revient en fait à simuler une réalisation d'un échantillon de N variables aléatoires $(X_i)_{0 \leq i < N}$ indépendantes de même loi que X.

11. D'après la Loi des grands nombres, si X est une variable d'espérance finie et si $(X_i)_{i \in \mathbb{N}}$

Est une suite de variables aléatoires indépendantes de même loi que X, alors la variable aléatoire $M_n = \sum_{0 \leq i < n} X_i$ converge presque sûrement vers la moyenne théorique $E(X)$ lorsque n tend vers $+\infty$.

11.1 Si l'entier N est assez grand, la moyenne des N réalisations simulées est donc assez proche de $E(X)$: $\frac{1}{N} \sum_{0 \leq i < N} x_i \approx E(X)$.

11.2 Si X est une variable aléatoire à valeurs dans N, alors la loi des grands nombres peut s'appliquer aux variables aléatoires de Bernoulli $[X_i = k]$ pour tout $k \in \mathbb{N}$, car il s'agit de variables aléatoires indépendantes, de même loi que $1[X = k]$.

On en déduit que, si l'entier N est assez grand, la fréquence empirique est assez proche de la probabilité théorique : $\# \{ 0 \leq i < N : x_i = k \} \approx N \cdot P(X = k)$.

12. De cette manière, la simulation d'un grand nombre de réalisations d'une variable aléatoire discrète X (au moyen d'un bon simulateur aléatoire) permet d'estimer la loi et l'espérance de cette variable.

12.1 Si la loi de X est connue, on peut ainsi tester la qualité du simulateur aléatoire.

12.2 Si la loi de X est difficile à calculer, on peut s'en faire une idée, soit en calculant des valeurs approchées des probabilités $P(X = k)$, soit en résumant les valeurs simulées sous la forme d'un histogramme.

Principe d'estimation d'une espérance à partir d'un échantillon

Soit (t_1, t_2, \dots, t_p) un échantillon de taille p de réalisations de la variable X .

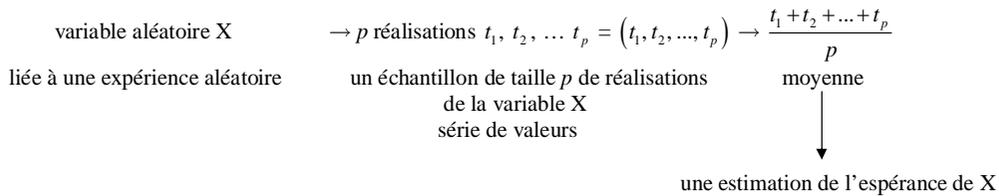
$\frac{t_1 + t_2 + \dots + t_p}{p}$ est une estimation de l'espérance de X .

Plus p est grand, plus cette estimation est précise.

Simulation

Programmes de simulation

Dans ce chapitre, nous allons reprendre l'interprétation de l'espérance d'une variable aléatoire.



Autrement dit, quand on calcule la moyenne des réalisations, on obtient une valeur plus ou moins proche de l'espérance.

L'estimation est d'autant meilleure que p est grand.

Cela nous donne un lien entre probabilités et statistiques.

Principe de calcul d'une moyenne avec Python

Simulation d'expériences aléatoires

Commencer par importer la bibliothèque correspondante : `>>> import numpy.random as rd`

5.1 Simulations d'expériences

Pour simuler la majorité des expériences aléatoires étudiées dans votre programme, il suffit de connaître deux commandes :

— `rd.randint(a,b)`, avec a et b des entiers : choisit au hasard un entier entre a et $b-1$.

Comme pour la syntaxe `range`, l'entier b est exclu.

— `rd.random()` (ne rien mettre entre les parenthèses !) : choisit au hasard un réel entre 0 et 1

5.2 Échantillons de lois usuelles

Pour obtenir des simulations de lois usuelles discrètes, voici les commandes à connaître :

— N simulations de $B(n, p)$: `rd.binomial(n,p,N)`

— N simulations de $U([a, b])$: `rd.randint(a,b,N)`

— N simulations de $G(p)$: `rd.geometric(p,N)`

— N simulations de $P(\lambda)$: `rd.poisson(lambda,N)`

Enfin, pour obtenir N réels entre 0 et 1 (l'an prochain, cette loi sera appelée loi uniforme sur $[0,1]$): `rd.random(N)`.

Pour faire des probabilités avec Python, on utilise la bibliothèque `numpy.random` que l'on importe à l'aide de la commande

1 `import numpy . random as rd`

1 La commande `rd.randint`

2 Simulation d'une épreuve de Bernoulli : commande `rd.binomial`

3 La commande `rd.random`

Deux syntaxes sont à connaître : `rd.random()` et `rd.randint(a,b)` cf. cours section 5.

La bibliothèque à importer est : `>>> import numpy.random as rd`

Exemple de programme Python permettant de simuler une variable aléatoire :

On lance un dé cubique non truqué dont les faces sont numérotées de 1 à 6.

On note le numéro de la face supérieure.

• Si le numéro obtenu est égal à 1, on gagne 12 €

• Si le numéro obtenu est égal à 2, on perd 6 €

• Si le numéro obtenu est égal à 3, 4, 5 ou 6, on ne gagne ni ne perd rien.

On note X la variable aléatoire donnant le gain algébrique du joueur en euros.

```
from random import randint
```

```
def gain():  
    r=randint(1,6)  
    if r==1:  
        x=12  
    elif r==2:  
        x=-6  
    else:  
        x=0  
    return x
```

```
def simulation(n):  
    s=0  
    for i in range(1,n+1):  
        s=s+gain()  
    return s/n
```

Le gain moyen est une approximation de l'espérance de X .

Fonction Python qui simule une réalisation d'une variable aléatoire X

sans argument



Fonction Python qui simule un échantillon de taille n de X
prend pour argument un entier naturel $n \geq 1$
renvoie une liste de taille n de réalisations de X

boucle for // liste



Fonction Python qui simule une réalisation de la variable aléatoire moyenne
 $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ où (X_1, X_2, \dots, X_n) est un échantillon de X

Modèle type pour un échantillon de taille n : $[X() \text{ for } _ \text{ in } \text{range}(n)]$

Voir document de Frédéric Junier

VI. Application au cas d'une loi binomiale

Dans ce paragraphe, nous allons donner une application historique importante de l'inégalité de Bienaymé-Tchebychev.

On considère une pièce truquée, telle que la probabilité d'obtenir pile en un lancer soit égale à p .
On effectue n lancers indépendants de cette pièce (n est un entier naturel non nul).
On note X_n la variable aléatoire égale au nombre de piles obtenus.

$$\text{Alors } \forall \varepsilon \in \mathbb{R}_+^* \quad P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2} \text{ et } \forall \varepsilon \in \mathbb{R}_+^* \quad P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

C'est Jacques Bernoulli (1654-1705) 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.

$\frac{X_n}{n}$ représente la fréquence de piles parmi les n lancers.

Démonstration :

X_n suit la loi binomiale de paramètres n (nombre d'épreuves) et p (probabilité d'obtenir pile en un lancer).

$$\text{On pose } Y_n = \frac{X_n}{n}.$$

On sait que l'espérance et la variance de X_n sont données par $E(X_n) = np$ et $V(X_n) = np(1-p)$.

$$E(Y_n) = \frac{E(X_n)}{n} = \frac{np}{n} = p$$

$$V(Y_n) = \frac{V(X_n)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

On applique l'inégalité de Bienaymé-Tchebychev à Y_n :

$$\forall \varepsilon \in \mathbb{R}_+^* \quad P\left(\left|Y_n - E(Y_n)\right| \geq \varepsilon\right) \leq \frac{V(Y_n)}{\varepsilon^2} \text{ qui donne immédiatement}$$

$$\forall \varepsilon \in \mathbb{R}_+^* \quad P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2} \quad (1).$$

Nous avons la première majoration.

Pour obtenir la seconde majoration, il suffit d'observer que le produit $p(1-p)$ est maximal pour $p = \frac{1}{2}$ et le maximum vaut $\frac{1}{4}$.

Il y a plusieurs moyens de justifier ce résultat. On peut, par exemple, considérer la fonction $f: x \mapsto x(1-x)$. C'est une fonction polynôme du second degré.

$$\text{On a donc } p(1-p) \leq \frac{1}{4}.$$

$$(1) \text{ donne donc } \forall \varepsilon > 0 \quad P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

Cette inégalité a une importance historique puisqu'elle apparaît dans les travaux de Jacques Bernoulli (meurt en 1705).

T spé

Exercices sur inégalités de Markov et de Bienaymé-Tchebychev

Inégalité de Markov

- 1** La moyenne des notes d'une classe à un devoir de mathématiques est de 12. À l'aide de l'inégalité de Markov, majorer la probabilité qu'un élève choisi au hasard dans cette classe ait une note supérieure ou égale à 14.
- 2** Dans un immeuble, l'ascenseur reste en moyenne 2 minutes au rez-de-chaussée avant d'être à nouveau sollicité. À l'aide de l'inégalité de Markov, majorer la probabilité que l'ascenseur reste au rez-de-chaussée plus de 5 minutes.
- 3** La température moyenne aux Maldives est de 28,4°C. On suppose que la température n'est jamais négative. À l'aide de l'inégalité de Markov, majorer la probabilité que la température soit supérieure ou égale à 34°C un jour donné.
- 4** Sur une autoroute la vitesse moyenne des voitures est de 120 km·h⁻¹.
1°) À l'aide de l'inégalité de Markov, majorer la probabilité qu'un automobiliste roule à une vitesse supérieure ou égale à 150 km·h⁻¹.
2°) À l'aide de l'inégalité de Markov, minorer la probabilité qu'un automobiliste roule à une vitesse strictement inférieure à 130 km·h⁻¹.

Inégalité de Bienaymé-Tchebychev

- 6** Sur les vingt matchs précédents, une équipe de rugby a marqué 60 essais. On note X la variable aléatoire donnant le nombre d'essais marqués au cours d'un match.
1°) Que vaut l'espérance de X ?
2°) On suppose que la variance de X est égale à 0,67. Majorer la probabilité qu'au cours du prochain match l'écart entre le nombre d'essais marqués et la moyenne soit supérieur ou égal à 1.
3°) Minorer la probabilité que l'écart entre le nombre d'essais marqués et la moyenne soit strictement inférieur à 2.
- 5** On lance 3600 fois une pièce de monnaie non truquée.
1°) À l'aide de l'inégalité de Bienaymé-Tchebychev, minorer la probabilité que le nombre d'apparitions de « Pile » soit strictement compris entre 1600 et 2000.
2°) Calculer la valeur exacte de cette probabilité.
- 7** Une classe présente les caractéristiques suivantes à un devoir : une moyenne de 12,4 et un écart-type de 1,2. Majorer la probabilité qu'un élève ait une note à ce devoir écartée d'au moins 2,5 points de la moyenne.
- 8** Dans une gare, le nombre moyen de passagers par jour est évalué à 5000 avec une variance de 2500. Majorer la probabilité que l'écart entre le nombre de visiteurs enregistré lors d'une journée et la moyenne soit supérieur ou égal à 100.
- 9** Des clients estiment que leur temps d'attente en magasin varie trop fortement selon les périodes. Le gérant décide donc de vérifier cette impression. Il mesure que le temps d'attente moyen de ses clients est de douze minutes et que la probabilité qu'un client attende strictement entre neuf et quinze minutes est de 0,55. Que peut-il en déduire sur la valeur minimale de l'écart type du temps d'attente de ses clients ?

Inégalité de concentration

- 10** 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.
- 11** On lance un dé tétraédrique équilibré dont les faces sont numérotées de 1 à 4. On note chaque fois le numéro de la face sur laquelle repose le dé. Combien de lancers doit-on effectuer pour être sûr au seuil de 99 % que la moyenne des résultats de ces lancers est comprise entre 2 et 3 exclus ?
- 12** Nicolas est facteur et distribue le courrier dans les logements d'un certain secteur géographique, ayant fait des études de mathématiques assez poussées, il a constaté que la probabilité d'un logement d'avoir du courrier un jour donné est égale à 0,6. Pour les besoins d'une enquête, Nicolas relève pendant 200 tournées supposées deux à deux indépendantes le nombre de logements pour lesquels il a distribué du courrier. Majorer la probabilité que le nombre de personnes ayant reçu du courrier ne soit pas dans l'intervalle $I =]100 ; 140[$.
- 13** Dans une société de démarchage par téléphone, on estime que 40 % des personnes appelées répondent effectivement. Lorsque l'on appelle 1000 personnes, majorer la probabilité que le nombre de personnes répondant soit en dehors de l'intervalle $I =]300 ; 500[$.
- 14** On considère un jeu dont le gain algébrique en euro est donné par une variable aléatoire X de loi de probabilité suivante.

x_i	-2	-1	5
$P(X = x_i)$	0,8	0,15	0,05

- 1°) Calculer l'espérance et la variance de X.
2°) On considère 1000 personnes ayant joué à ce jeu durant une semaine. Minorer la probabilité que le gain algébrique moyen sur ces 1000 personnes soit dans l'intervalle $I =]-1,7 ; -1,3[$.

15 On considère une marche aléatoire sur \mathbb{Z} définie de la façon suivante : on part de 0 et, à chaque étape, indépendamment, on a une probabilité $\frac{1}{2}$ de faire un pas vers la droite et une probabilité $\frac{1}{2}$ de faire un pas vers la gauche.

Pour tout entier naturel i non nul, on note X_i la variable aléatoire dont la loi de probabilité est donnée par

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}.$$

Pour tout entier naturel non nul n , on pose $S_n = X_1 + \dots + X_n$.

On représente une « trajectoire » par une ligne brisée joignant les points de coordonnées $(n, S_n(\omega))$.

Les valeurs de n sont portées en abscisse.

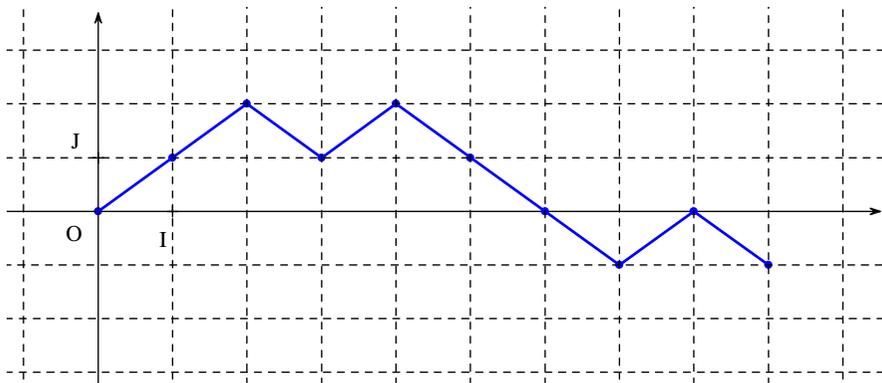
Voici un exemple de valeurs de X_i :

i	0	1	2	3	4	5	6	7	8	9
Valeur de X_i	0	1	1	-1	1	-1	-1	-1	1	-1

X_0 prend initialement la valeur 0 puisque l'on part de 0.

On peut calculer S_0, S_1, S_2 etc.

On peut représenter la « trajectoire » sur un graphique par une ligne brisée.



1°) Que représente S_1 ? S_2 ? S_n pour n entier naturel quelconque non nul ?

Il s'agit de sommes.

2°) Déterminer l'espérance et la variance X_i pour i entier naturel supérieur ou égal à 1.

En déduire l'espérance et la variance de S_n .

3°) Déterminer pour x réel strictement positif, une majoration de $P\left(\left|\frac{S_n}{n}\right| \geq x\right)$.

Loi faible des grands nombres et simulations

Rappel des deux fonctions de la bibliothèque random utilisables pour le programme :

- La fonction random renvoie un réel au hasard de l'intervalle $[0; 1[$.

- La fonction uni form(a, b) renvoie un réel au hasard dans l'intervalle $[a; b]$ (les arguments sont deux réels a et b tels que $a \leq b$).

On privilégiera la fonction uni form.

16 Pour une mise de 3 € un joueur peut lancer un dé cubique équilibré dont les faces sont numérotées de 1 à 6.

Avec un 6, il gagne 2 €

Avec un résultat compris 3 et 5 tous les deux inclus, il gagne 1 €

Avec un résultat inférieur ou égal à 2, il ne gagne rien.

Le joueur doit ensuite lancer une pièce équilibrée. Si la face visible est pile, il multiplie ses gains par 3 ; sinon, il les multiplie par 4.

1°) Écrire un programme Python de simulation permettant de calculer la moyenne du gain pour un échantillon de taille n .

2°) Réaliser ce programme.

En déduire une estimation de l'espérance de gain, mise déduite.

Comparer avec la valeur exacte de l'espérance.

17 En présence de trois joueurs A, B, C, on lance trois dés cubiques équilibrés (un rouge, un bleu et un jaune) dont les faces sont numérotées de 1 à 6.

Le joueur A gagne 1 € si le résultat du dé rouge est inférieur à celui des deux autres dés.

Le joueur B gagne 1 € si le résultat du dé bleu est supérieur à celui des deux autres dés.

Le joueur C gagne 1 € si le résultat du dé jaune est compris strictement entre celui des deux autres dés.

Dans les autres cas, personne ne gagne.

1°) Écrire une fonction Python qui renvoie sous la forme d'une liste le montant en euros des trois joueurs après n lancers de dés.

2°) À l'aide d'un échantillon de 10000 parties, donner une estimation du gain moyen.

18 Une urne contient quatre boules rouges et une boule bleue. On réalise jusqu'à 10 tirages avec remise dans cette urne. Les tirages s'arrêtent dès que la boule bleue est tirée et on note G le nombre de boules rouges tirées.

1°) Recopier et compléter la fonction Python d'en-tête def si mu_G : suivante afin de simuler la variable aléatoire G .

```

from random import randint

def si mu_G() :
    g=0
    for tirage in range(10) :
        if randint(1,5)..... :
            return .....
        g=g+1
    return g

```

On décide de numéroter les boules : 1, 2, 3, 4 pour les boules rouges et 5 la boule bleue.

2°) Coder une fonction Python d'en-tête def ech(N) : qui renvoie une liste de N réalisations de la variable aléatoire moyenne M_n de la loi de probabilité G .

3°) La probabilité de tirer la boule bleue en un lancer est $p = 0,2$ et la probabilité de tirer une boule rouge en un lancer est $q = 0,8$.

Vérifier que la loi de probabilité de G est donnée dans le tableau ci-dessous.

g_i	0	1	2	3	4	5	6	7	8	9	10
$P(G = g_i)$	p	pq	pq^2	pq^3	pq^4	pq^5	pq^6	pq^7	pq^8	pq^9	q^{10}

Calculer l'espérance de G .

Vérifier que les résultats obtenus par simulation sont proches de cette valeur.

18 bis Une urne contient une boule blanche et des boules noires. On réalise jusqu'à n tirages avec remise dans cette urne. Les tirages s'arrêtent dès que la boule blanche est tirée et on note X le nombre de boules noires tirées.

On note p la probabilité de tirer la boule blanche et q la probabilité de tirer une boule noire en un lancer.

1°) Vérifier que la loi de probabilité de X est donnée dans le tableau.

x_i	0	1	2				$n-1$	n
$P(X = x_i)$	p	pq	pq^2					pq^{n-1}	q^n

2°) On admet que $E(X) = \frac{q}{p}(1 - q^n)$.

Calculer la valeur arrondie au millième de $E(X)$ lue l'urne contient 4 boules noires (et toujours une boule blanche).

19 On souhaite estimer le nombre de lancers nécessaires pour faire apparaître toutes les faces d'un dé cubique équilibré dont les faces sont numérotées de 1 à 6. Pour cela, on utilise l'algorithme suivant :

```

a, b, c, d, e, f ← 0, 0, 0, 0, 0, 0
N ← 0
Tantque abcdef = 0 Faire
    z ← entier aléatoire entre 1 et 6
    Si z = 1
        a ← 1
    FinSi
    Si z = 2
        b ← 1
    FinSi
    Si z = 3
        c ← 1
    FinSi
    Si z = 4
        d ← 1
    FinSi
    Si z = 5
        e ← 1
    FinSi
    Si z = 6
        f ← 1
    FinSi
    N ← N + 1
FinTantque

```

1°) Expliquer le fonctionnement de cet algorithme.

2°) Modifier cet algorithme pour qu'il simule n fois l'expérience qui consiste à compter le nombre de lancers nécessaires pour que toutes les faces soient sorties une fois, et qui calcule ensuite le nombre moyen de lancers nécessaires.

3°) Programmer cet algorithme en Python. En déduire une estimation du nombre moyen de lancers nécessaires pour faire apparaître toutes les faces du dé.

4°) On peut démontrer que la valeur théorique est $6 \times \left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right) = 14,7$.

Vérifier que les résultats obtenus par simulation sont proches de cette valeur.

C'est le problème du collectionneur.

20 On choisit au hasard réel quelconque entre 0 et 1, et on répète cela jusqu'à ce que la somme des nombres choisis dépasse 1.

Exemple avec des nombres décimaux (théoriquement, on choisit des réels, pas forcément décimaux) :

1^{er} nombre choisi : 0,372044

Comme il n'est pas supérieur à 1, on choisit un nouveau nombre au hasard entre 0 et 1.

2^e nombre choisi : 0,613

Comme la somme des deux premiers nombres choisis 0,372044 + 0,613 est inférieure ou égale à 1, on choisit un nouveau nombre au hasard entre 0 et 1.

3^e nombre choisi : 0,001007

Comme la somme des trois premiers nombres choisis 0,372044 + 0,613 + 0,001007 est inférieure ou égale à 1, on choisit un nouveau nombre au hasard entre 0 et 1.

4^e nombre choisi : 0,828

Cette fois, la somme des quatre premiers nombres choisis 0,372044 + 0,613 + 0,001007 + 0,828 dépasse 1.

On s'arrête là.

Dans cet exemple, il a fallu choisir 4 nombres au hasard entre 0 et 1 pour que la somme dépasse 1.

Conjecturer à l'aide d'un programme Python la valeur exacte du nombre moyen de nombres choisis.

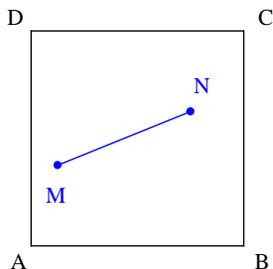
Ce problème est parfois désigné comme « paradoxe de la grenouille » ou « paradoxe des grenouilles » en référence à la formulation suivante du problème :

« Une grenouille veut rejoindre une mare située à un mètre devant elle. Pour l'atteindre, elle effectue des sauts supposés en ligne droite de longueur un nombre aléatoire de l'intervalle $[0; 1[$. »

21 Soit ABCD un carré de côté 1.

Soit M et N deux points au hasard situés à l'intérieur du carré.

On s'intéresse à la distance MN.



On peut noter d'emblée que la distance maximale de deux points quelconques situés à l'intérieur du carré est $\sqrt{2}$.

1°) Le but de cette question est de réaliser un programme Python permettant d'obtenir une estimation de la valeur moyenne de cette distance.

On se place dans un repère orthonormé (O, \vec{i}, \vec{j}) tel que O soit confondu avec A, $\overline{AB} = \vec{i}$ et $\overline{AD} = \vec{j}$.

a) Caractériser l'intérieur du carré ABCD par un système d'inéquations.

b) Rédiger le programme Python correspondant.

Rappel des deux fonctions de la bibliothèque random utilisables pour le programme :

- La fonction random renvoie un réel au hasard de l'intervalle $[0; 1[$;

- La fonction uniform(a, b) renvoie un réel au hasard dans l'intervalle $[a; b]$ (les arguments sont deux réels a et b tels que $a \leq b$).

On privilégiera la fonction uniform.

c) On peut démontrer que la distance moyenne est égale à $\frac{2 + \sqrt{2} + 5 \ln(\sqrt{2} + 1)}{15}$.

Comparer les estimations obtenues par simulation avec la valeur exacte.

3°) Faire le même travail pour deux points aléatoires :

- dans un cube d'arête 1 ;

- dans un disque de rayon 1.

22 On effectue une suite de lancers d'une pièce de monnaie. On suppose que les résultats des lancers sont indépendants et que, à chaque lancer, la pièce donne pile avec la probabilité p ($0 < p < 1$) et face avec la probabilité $q = 1 - p$.

1°) Réaliser un programme Python permettant d'estimer le nombre de lancers nécessaires pour obtenir deux piles de suite, c'est-à-dire lors de deux lancers consécutifs.

2°) Donner une estimation dans le cas où $p = \frac{1}{2}$ (cas d'une pièce équilibrée) et dans le cas où $p = \frac{1}{3}$.

3°) On peut démontrer que la valeur théorique est $\frac{1+p}{p^2}$.

Dans les deux cas, comparer les résultats obtenus par simulation avec la valeur théorique.

23 Soit A et B deux points quelconques du plan tels que $AB = 1$.

Soit M et N deux points au hasard appartenant au segment $[AB]$.

On s'intéresse à la distance MN.

Réaliser un programme Python permettant d'obtenir une estimation de la valeur moyenne de cette distance.

Conjecturer la valeur exacte de cette distance moyenne.

Propriété :

Une unité de longueur est fixée.

Soit Δ un axe de repère (O, \vec{i}) tel que $\|\vec{i}\| = 1$.

Soit M et N deux points de Δ d'abscisses respectives x et y.

La distance MN est égale à la valeur absolue de $x - y$.

24 On dispose d'une pièce non truquée. On étudie deux jeux.

Premier jeu :

On lance la pièce jusqu'à l'apparition du premier « Pile ».

On note X la variable aléatoire égale au nombre de fois où l'on a obtenu « Face ».

Rédiger une fonction Python permettant de simuler X.

Déterminer une estimation de l'espérance de X.

On admet que l'espérance de X est égale à $\frac{1}{p} - 1$.

Deuxième jeu :

On lance la pièce jusqu'à l'apparition du second « Pile ».

On note Y la variable aléatoire égale au nombre de fois où l'on a obtenu « Face ».

Rédiger une fonction Python permettant de simuler Y.

Déterminer une estimation de l'espérance de Y.

On admet que l'espérance de Y est égale à $\frac{2}{p} - 2$.

Reprendre les questions précédentes en supposant que la pièce est truquée.

25 On fait tourner n fois une roulette à deux secteurs, l'un marqué 1 et l'autre 0. La probabilité d'obtenir le secteur 1 est notée p .

Dans la suite de 1 et de 0 obtenus, on s'intéresse aux « blocs » de 0 et 1. Par exemple, la suite 10010110101 contient 9 blocs : quatre blocs de 0 et cinq blocs de 1.

À chaque suite de n chiffres obtenus, la variable aléatoire S associe le nombre de blocs de la suite.

Par exemple, pour la suite définie plus haut, S prend la valeur $4 + 5 = 9$.

Pour tout entier naturel i tel que $1 \leq i \leq n$, on appelle X_i la variable aléatoire qui, au i -ième tour de roulette, associe le numéro du secteur obtenu.

Pour tout entier naturel i tel que $2 \leq i \leq n$, on appelle Y_i la variable aléatoire telle que $Y_i = 1$ si X_{i-1} est différent de X_i et 0 sinon.

1°) a) Donner les valeurs de $P(X_i = 0)$ et de $P(X_i = 1)$.

b) Justifier que : $P(Y_i = 1) = P(X_{i-1} = 0) \times P(X_i = 1) + P(X_{i-1} = 1) \times P(X_i = 0)$.

c) Déterminer $P(Y_i = 1)$ en fonction de p .

2°) a) Exprimer S en fonction de Y_2, Y_3, \dots, Y_n puis en déduire l'espérance de S en fonction de n et p .

b) Quel nombre moyen de blocs obtient-on pour $n = 1001$ et $p = 0,5$?

3°) On donne le script incomplet d'une fonction Python qui retourne le nombre de blocs dans le cas $n = 1001$ et $p = 0,5$ de la suite définie à la ligne `L=[... for k in range(1001)]`.

On crée plusieurs fonctions.

Première fonction : compte le nombre de blocs dans une liste constituée de 0 et de 1 à compléter :
La liste contient au moins deux éléments.

```
def blocs01(L):
    t=(L[1]!=L[0])
    for i in range(1, len(L)):
        t=t+...
    return (t+1)
```

26 Pierre Rémond de Montmort rapporte dans son *Essai d'analyse sur les jeux de hasard* (1708) un jeu observé par un explorateur revenant du Canada.

On y joue avec 8 noyaux noirs d'un côté et blanc de l'autre ; on jette les noyaux en l'air : alors, si les noyaux noirs se trouvent impairs, celui qui a jeté les noyaux gagne ce que l'autre a mis au jeu.

S'ils se trouvent ou tous noirs ou tous blancs, il en gagne le double ; et hors de ces deux cas, il perd sa mise.

Deux amis A et B jouent.

Le joueur A lance les noyaux.

```
from random import random

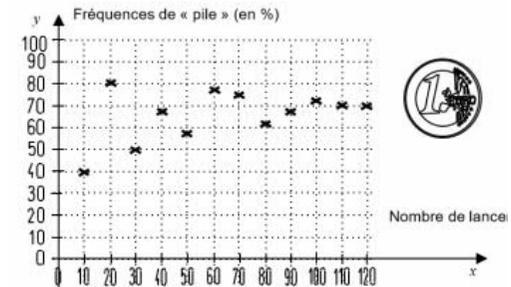
def lancer():
    L=[]
    for i in range(8) :
        if random()<0.5:
            L.append('N')
        else:
            L.append('B')
    return L

def partie(mise):
    gain=-mise
    L=lancer()
    N=L.count('N')
    if N%2==1:
        gain=gain+mise
    elif N==0 or N==8:
        gain=gain+2*mise
    return
```

27 Les fréquences d'obtention de « Pile » lors de séries indépendantes de lancers d'une pièce « truquée » sont représentées sur le graphique ci-dessous.

Lorsque le nombre de lancers augmente, les fréquences se stabilisent.

Quelle est la probabilité d'obtenir « Pile » avec cette pièce ?



Solutions

1

On note X la note d'un élève de la classe pris au hasard.

Comme la moyenne de la classe en mathématiques est 12, l'espérance de X est 12 (c'est-à-dire $E(X) = 12$).

X est une variable aléatoire dont les valeurs sont positives ou nulles.

D'après l'inégalité de Markov, on a $P(X \geq 14) \leq \frac{12}{14}$ soit $P(X \geq 14) \leq \frac{6}{7}$.

La calculatrice donne $\frac{6}{7} = 0,8571428\dots$

2

On note X le temps en minutes où l'ascenseur reste au rez-de-chaussée.

Comme l'ascenseur reste en moyenne 2 minutes au rez-de-chaussée avant d'être à nouveau sollicité, on en déduit que $E(X) = 2$.

X est une variable aléatoire dont les valeurs sont positives ou nulles.

D'après l'inégalité de Markov, on a $P(X \geq 5) \leq \frac{2}{5}$ soit $P(X \geq 5) \leq 0,4$.

3

On note X la température aux Maldives en °C.

Comme la température moyenne est de 28,4°C, on en déduit que $E(X) = 28,4$.

X est une variable aléatoire dont les valeurs sont positives ou nulles (vu la latitude des Maldives, on peut dire que la température y est toujours positive ou nulle).

D'après l'inégalité de Markov, on a $P(X \geq 34) \leq \frac{28,4}{34}$, soit $P(X \geq 34) \leq \frac{71}{85}$.

Il est possible de simplifier $\frac{28,4}{34}$ en $\frac{14,2}{17}$.

4

Soit X la vitesse d'un automobiliste en $\text{km} \cdot \text{h}^{-1}$.

Comme la vitesse moyenne des voitures est de 120 $\text{km} \cdot \text{h}^{-1}$, on en déduit que $E(X) = 120$.

X est une variable aléatoire dont les valeurs sont positives ou nulles.

1°)

D'après l'inégalité de Markov, $P(X \geq 150) \leq \frac{120}{150}$, soit $P(X \geq 150) \leq \frac{4}{5}$.

2°)

On écrit $P(X < 130) = 1 - P(X \geq 130)$.

D'après l'inégalité de Markov, on a $P(X \geq 130) \leq \frac{120}{130}$, soit $P(X \geq 130) \leq \frac{12}{13}$.

En multipliant les deux membres de cette inégalité par -1 , on obtient $-P(X \geq 130) \geq -\frac{12}{13}$.

En ajoutant ensuite 1 à chaque membre de cette inégalité, on obtient $1 - P(X \geq 130) \geq \frac{1}{13}$ soit

$$P(X < 130) \geq \frac{1}{13}.$$

5

1°) Soit X la variable aléatoire qui compte le nombre de piles obtenus à l'issue des 3600 lancers.

X suit la loi binomiale de paramètres $n = 3600$ (nombre de répétitions) et $p = \frac{1}{2}$ (probabilité d'un succès).

On a $E(X) = 3600 \times \frac{1}{2} = 1800$ et $V(X) = 3600 \times \frac{1}{2} \times \frac{1}{2} = 900$.

On utilise les formules donnant l'espérance et la variance d'une variable aléatoire X qui suit la loi binomiale de paramètres n (nombre de répétitions) et p (probabilité d'un succès) rappelées dans l'encadré ci-dessous :

$$\begin{aligned} E(X) &= n \times p \\ V(X) &= n \times p \times (1 - p) \end{aligned}$$

On note A l'événement « Le nombre d'apparitions de piles est strictement compris entre 1600 et 2000 ».

$$A = (1600 < X < 2000) = (|X - 1800| < 200)$$

On a $P(A) = 1 - P(|X - 1800| \geq 200)$.

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X - 1800| \geq 200) \leq \frac{900}{200^2}$, soit

$$P(|X - 1800| \geq 200) \leq 0,0225.$$

On a donc $-P(|X - 1800| \geq 200) \geq -0,0225$.

En ajoutant 1 à chaque membre de cette inégalité, on obtient $P(A) \geq 0,9775$.

$$2^\circ) P(A) = P(1601 \leq X \leq 1999)$$

X suit la loi binomiale de paramètres $n = 3600$ (nombre de répétitions) et $p = \frac{1}{2}$ (probabilité d'un succès).

Avec la calculatrice, on obtient l'affichage 1.

Cela signifie que $P(A)$ est un nombre décimal très proche de 1.

6

$$1^\circ) E(X) = \frac{60}{20} = 3$$

$$2^\circ) V(X) = 0,67$$

D'après l'inégalité de Bienaymé-Tchebychev, on a $P(|X-3| \geq 1) \leq \frac{0,67}{1^2}$ soit $P(|X-3| \geq 1) \leq 0,67$.

3°)

$$\text{On a } P(|X-3| < 2) = 1 - P(|X-3| \geq 2).$$

D'après l'inégalité de Bienaymé-Tchebychev, on a $P(|X-3| \geq 2) \leq \frac{0,67}{4}$.

Donc $P(|X-3| < 2) \geq 1 - \frac{0,67}{4}$ soit $P(|X-3| < 2) \geq 0,8325$.

7

Soit X la variable aléatoire égale à la note d'un élève au devoir.

D'après l'énoncé, on a $E(X) = 12,4$.

$$\sigma(X) = 1,2 \text{ donc } V(X) = [\sigma(X)]^2 = 1,2^2 = 1,44.$$

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X-12,4| \geq 2,5) \leq \frac{1,44}{2,5^2}$ soit

$$P(|X-12,4| \geq 2,5) \leq 0,2304.$$

8

On note X le nombre de passagers par jour dans la gare.

D'après l'énoncé, on a $E(X) = 5000$ et $V(X) = 2500$.

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X-5000| \geq 100) \leq \frac{2500}{100^2}$, soit $P(|X-5000| \geq 100) \leq \frac{1}{4}$.

9

On note X la variable aléatoire qui compte le temps d'attente d'un client en minutes.

On a $E(X) = 12$.

On a $9 < X < 15 \Leftrightarrow |X-12| < 3$ (centre de l'intervalle : 12 ; rayon : 3).

On a $P(9 < X < 15) = 0,55$ donc $P(|X-12| \geq 3) = 0,45$.

D'après l'inégalité de Bienaymé-Tchebychev, $P(|X-12| \geq 3) \leq \frac{[\sigma(X)]^2}{3^2}$ donc

$$[\sigma(X)]^2 \geq 3^2 \times P(|X-12| \geq 3)$$

On en déduit que $\sigma(X) \geq 3\sqrt{P(|X-12| \geq 3)}$ soit $\sigma(X) \geq 3\sqrt{0,45}$

Avec la calculatrice, on obtient $3\sqrt{0,45} = 2,0124\dots$

10

On numérote les personnes de 1 à 100.

On note X_i le gain en euros de la personne numéro i .

$(X_1, X_2, \dots, X_{100})$ est un échantillon de variables aléatoires d'espérance 10 et de variance 2.

On note M la variable aléatoire moyenne de cet échantillon.

On cherche à minorer la probabilité que $7 < M < 13$ soit $|M-10| < 3$ (on peut écrire $7 = 10 - 3$ et $13 = 10 + 3$).

On passe par l'événement contraire.

L'inégalité de concentration vue dans la propriété déduite de l'inégalité de Bienaymé-Tchebychev (inégalité de concentration) donne :

$$P(|M-10| \geq 3) \leq \frac{2}{100 \times 9} \text{ soit } P(|M-10| \geq 3) \leq \frac{1}{450}.$$

On obtient $P(|M-10| < 3) \geq \frac{449}{450}$.

11

On note n le nombre de lancers.

On note X_i la variable aléatoire donnant le résultat du lancer numéro i .

(X_1, X_2, \dots, X_n) est un échantillon de variables aléatoires d'espérance $\frac{5}{2}$ et de variance $\frac{5}{4}$ (on peut utiliser directement les résultats donnant l'espérance et la variance d'une variable aléatoire qui suit une loi uniforme).

On note M la variable aléatoire moyenne de cet échantillon.

On cherche n tel que $P(2 < M < 3) \geq 0,99$ (1).

$$(1) \Leftrightarrow P(|M - 2,5| < 0,5) \geq 0,99$$

$$\Leftrightarrow P(|M - 2,5| \geq 0,5) \leq 0,01$$

Or par l'inégalité de concentration déduite de l'inégalité de Bienaymé-Tchebychev,

$$P(|M - 2,5| \geq 0,5) \leq \frac{\frac{5}{n}}{n \times \frac{1}{4}} \text{ soit } P(|M - 2,5| \geq 0,5) \leq \frac{5}{n}.$$

Donc on est assuré que (1) est vérifiée dès lors que $\frac{5}{n} \leq 0,01$ soit $n \geq 500$.

On a effectué un raisonnement par condition suffisante.

Il suffit de choisir n supérieur ou égal à 500 pour garantir que (1) soit vérifiée.

12

On note X la variable aléatoire qui compte le nombre de logements pour lesquels Nicolas a déposé du courrier. X suit la loi binomiale de paramètres $n = 200$ (nombre de répétitions) et $p = 0,6$ (probabilité d'un succès).

On a $E(X) = 120$ et $V(X) = 48$.

On cherche à majorer la probabilité $P(X \notin I)$.

On observe que $100 = 120 - 20$ et $140 = 120 + 20$.

D'après l'inégalité de Bienaymé-Tchebychev, on a $P(X \notin I) = P(|X - 120| \geq 20) \leq \frac{48}{20^2}$ soit $P(X \notin I) \leq \frac{3}{25}$.

Il est possible de calculer $P(X \in I)$ et $P(X \notin I)$ avec la calculatrice.

Pour commencer, on peut observer que l'ensemble des entiers naturels appartenant à I est l'intervalle d'entiers $[[101; 139]]$.

Pour pouvoir utiliser la fonction de répartition de la loi binomiale, on écrit

$$P(X \in I) = P(X \leq 139) - P(X \leq 100).$$

On obtient l'affichage : 0,9952343369.

On en déduit que $P(X \notin I) = 0,004765663...$

Or on a $\frac{3}{25} = 0,12$ donc le résultat est bien en accord avec la majoration $P(X \notin I) \leq \frac{3}{25}$.

13

On note X la variable aléatoire qui compte le nombre de personnes répondant.

X suit la loi binomiale de paramètres $n = 1000$ (nombre de répétitions) et $p = 0,4$ (probabilité qu'une personne réponde).

On a $E(X) = 400$ et $V(X) = 240$.

On cherche la probabilité $P(X \notin I)$.

On observe que $300 = 400 - 100$ et $500 = 400 + 100$.

D'après l'inégalité de Bienaymé-Tchebychev, on a $P(X \notin I) = P(|X - 400| \geq 100) \leq \frac{240}{100^2}$ soit

$$P(X \notin I) \leq 0,024 \text{ (ou } P(X \notin I) \leq \frac{3}{125}).$$

Il est possible de calculer $P(X \in I)$ et $P(X \notin I)$ avec la calculatrice.

Pour commencer, on peut observer que l'ensemble des entiers naturels appartenant à I est l'intervalle d'entiers $[[301; 499]]$.

X suit la loi binomiale de paramètres $n = 1000$ et $p = 0,4$.

Pour pouvoir utiliser la fonction de répartition de la loi binomiale, on écrit

$$P(X \in I) = P(X \leq 499) - P(X \leq 300).$$

On obtient l'affichage : 0,999999999999.

Pour $P(X \notin I)$, on obtient l'affichage : $1,3298 \times 10^{-10}$.

Le résultat est donc bien en accord avec la majoration $P(X \notin I) \leq 0,024$.

14

1°)

$$E(X) = -2 \times 0,8 - 1 \times 0,15 + 5 \times 0,05 \\ = -1,5$$

$$V(X) = (-2)^2 \times 0,8 + (-1)^2 \times 0,15 + 5^2 \times 0,05 - (-1,5)^2 \\ = 2,35$$

2°) Les 1000 personnes constituent un échantillon $(X_1, X_2, \dots, X_{1000})$ de X .

On note M la moyenne de cet échantillon.

On cherche à minorer $P(M \in I)$.

On observe que $-1,7 = -1,5 - 0,2$ et $-1,3 = -1,5 + 0,2$.

$$M \in I \Leftrightarrow |M + 1,5| < 0,2$$

On utilise l'événement contraire $(M \notin I)$.

$$M \notin I \Leftrightarrow |M + 1,5| \geq 0,2$$

L'inégalité de concentration donne $P(M \notin I) \leq \frac{2,35}{1000 \times 0,2^2}$ soit $P(M \notin I) \leq \frac{2,35}{40}$ ce qui donne finalement

$$P(M \notin I) \leq 0,05875.$$

On en déduit facilement $P(M \in I) \geq 0,94125$.

15

On peut faire un graphique.

- 1°)
 S_1 représente l'abscisse du point en lequel on se trouve après 1 étape.
 S_2 représente l'abscisse du point en lequel on se trouve après 2 étapes.

S_n représente l'abscisse du point en lequel on se trouve après n étapes.

2°)

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

$$E(X_i) = 0$$

$$V(X_i) = E((X_i - E(X_i))^2) = E((X_i - 0)^2) = E(X_i^2) = 1^2 \times \frac{1}{2} + (-1)^2 \times \frac{1}{2} = 1$$

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = 1^2 \times \frac{1}{2} + (-1)^2 \times \frac{1}{2} - 0^2 = 1$$

3°)

$$E(S_n) = \sum_{k=1}^{k=n} E(X_i) = 0$$

Les variables X_1, X_2, \dots, X_n sont mutuellement indépendantes [on dit que (X_1, X_2, \dots, X_n) est un échantillon de variables de même loi].

On peut donc appliquer la propriété sur la variance d'une somme.

$$V(S_n) = \sum_{k=1}^{k=n} V(X_i) = \sum_{k=1}^{k=n} 1 = n$$

4°) On écrit $P\left(\left|\frac{S_n}{n}\right| \geq x\right) = P(|S_n| \geq nx) = P(|S_n - E(S_n)| \geq nx)$.

On applique ensuite l'inégalité de Bienaymé-Tchebychev.

$P(|S_n - E(S_n)| \geq nx) \leq \frac{V(S_n)}{(nx)^2}$ soit $P(|S_n| \geq nx) \leq \frac{n}{n^2 x^2}$ ce qui donne finalement $P(|S_n| \geq nx) \leq \frac{1}{nx^2}$.

16

1°)

```
from random import randint

def de() :
    r=randint(1 ;6)
    if r<=2
        return 0
    elif r>5 :
        return 2
    else :
        return 1

def piece() :
    s= randint(0,1)
    if g=0 :
        return 3
    else :
        return 9

def gain(n) :
    L= [pièce ()* dé () - 3 for i in range (n)]
    return sum(L)/n
```

2°)

On définit les événements suivants.

- A : « obtenir un numéro inférieur ou égal à 2 »
- B : « obtenir un numéro entre 3 et 5 »
- C : « obtenir un 6 »
- D : « obtenir pile »
- E : « obtenir face »

On dresse un tableau de probabilités.

x_i	-3	0	1	3	5	
$P(X=x_i)$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	1

$$E(X) = -1 + \frac{1}{4} + \frac{1}{4} + \frac{5}{12}$$

$$= -\frac{1}{12}$$

17

```

from random import randint

def jeu() :
    a, b, c=0, 0, 0
    L=[randint(1,6) for i in range (2)]
    if L[1]<L[2] and L[1]<L[0]:
        a+=1
    if L[0]>L[1] and L[0]>L[2]:
        b+=1
    if L[2]<L[0] and L[2]>L[1] or L[2]<L[1] and L[2]>L[0]:
        c+=1
    return a+b+c

def gain(n):
    G=[jeu() for i in range(n)]
    return sum(G)/n

```

18

1°)

```

from random import randint

def simu_G() :
    g=0
    for tirage in range(10) :
        if randint(1,5)..... :
            return .....
            g=g+1
    return g

```

Le programme suivant est faux.

```

from random import randint

def simu_G() :
    g=0
    while rand(1,5)!=5 :
        g=g+1
    return g

```

2°) Coder une fonction ech qui renvoie une liste de N réalisations de la variable aléatoire moyenne M_n de la loi de probabilité G.

3°) Il s'agit d'une variable aléatoire apparentée à une loi géométrique tronquée.

$$E(G) = 0 \times p + 1 \times pq + 2pq^2 + 3pq^3 + \dots + 9pq^9 + 10q^{10}$$

$$= p(q + 2q^2 + 3q^3 + \dots + 9q^9) + 10q^{10}$$

$$= \frac{34868196}{9765625}$$

$$= 3,5705032\dots \text{ (nombre décimal)}$$

$$5(1 - 3 \times 0,8^{10}) = 3,38938\dots$$

On peut aussi utiliser la commande somme de la calculatrice.

$$E(G) = \sum_{k=0}^{k=n} (k \times 0,2 \times 0,8^k) + 10 \times 0,8^{10}$$

$$18 \quad \sum_{k=1}^n kx^k = \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(1-x)^2}$$

Le 7-6-2023

$$k \times 0,2 \times (0,8)^k + 10 \times 0,8^{10}$$

$$\frac{1}{p} [1 - (1+np)(1-p)^n]$$

Loi géométrique tronquée

$$18 \text{ bis} \quad \sum_{k=0}^{k=n} kx^{k-1} = \frac{-(n+1)x^n(1-x) - (1-x^{n+1}) \times (-1)}{(1-x)^2}$$

$$\sum_{k=0}^{k=n} kx^{k-1} = \frac{1 - x^{n+1} - (n+1)x^n + (n+1)x^n}{(1-x)^2}$$

$$\sum_{k=0}^{k=n} kx^{k-1} = \frac{nx^{n+1} - (n+1)x^n + 1}{(1-x)^2}$$

$$E(X) = \sum_{k=0}^{k=n-1} kpq^k + n \times q^n$$

$$E(X) = p \sum_{k=0}^{k=n-1} kq^k + n \times q^n$$

$$E(X) = p \times \frac{q[(n-1)q^n - nq^{n-1} + 1]}{p^2} + nq^n$$

$$E(X) = \frac{q}{p} \times [(n-1)q^n - nq^{n-1} + 1] + nq^n$$

$$E(X) = \frac{1}{p} \times [(n-1)q^{n+1} - nq^n + q + npq^n]$$

$$E(X) = \frac{1}{p} \times [(n-1)q^{n+1} - nq^{n+1} + q]$$

$$E(X) = \frac{1}{p} \times (q - q^{n+1})$$

$$E(X) = \frac{q}{p} (1 - q^n)$$

$$\frac{0,8}{0,2} (1 - 0,8^{10}) = 4(1 - 0,8^{10}) = 3,5705032704... \text{ (nombre décimal)}$$

19

Le 14-6-2023 (ou 5-6-2023)

Problème du collectionneur

```
from random import randint
```

```
L=10*[0]
n=0
i=0
while i<10 :
    n=n+1
    x=randint(0, 9)
    if L[x]==0 :
        L[x]=1
        i=i+1
    print (L)
print (n)
```

20

Le 3-6-2023

```
from random import random
```

```
def grenouille():
    s=0
    c=0
    while s<=1:
        s=s+random()
        c=c+1
    return c

def estim_nbre_moy(n):
    L=[grenouille() for _ in range(n)]
    return sum(L)/n
```

On s'intéresse au nombre de nombres choisis.

Le problème peut s'implémenter en Python à l'aide de la fonction suivante :

```
from random import random
```

```
def moyenne(n):
    T = 0
    for k in range(n):
        s, i = 0, 0
        while s < 1:
            s += random()
            i += 1
        T += i
    return T / n
```

Autre proposition élève avec une liste le 25-5-2022 :

```

from random import uniform

def moyenne(n):
    L = []
    for i in range(n):
        S = 0
        N=0
        while S < 1:
            r=uniform(0, 1)
            S += r
            N+= 1
        L.append(N)
    return sum(L) / len(L)

```

On obtient alors :

```

>>> moyenne(10**7)
2. 7186881
>>> moyenne(10**8)
2. 71829368
>>> moyenne(10**9)
2. 718286834

```

La moyenne semble être égale à e (nombre de Néper). Ce n'est qu'une conjecture.

On peut démontrer cette conjecture grâce à des outils qui dépassent le programme de terminale.

Ce problème s'appelle aussi « paradoxe des grenouilles » ou « paradoxe de la grenouille ».

Le 3-6-2023

J'avais écrit le programme suivant sur la même feuille en dessous du programme grenouille().

```

def échant(n):
    L=[grenouille() for _ in range(n)]
    return L

```

21

Dans cet exercice, on cherche à estimer la distance moyenne entre deux points pris au hasard à l'intérieur d'un carré de côté 1.

1°)

$$a) \begin{cases} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{cases}$$

```

from random import random

def dist():
    # choisit 2 points au hasard et retourne la distance entre eux
    x1 = random()
    x2 = random()
    y1 = random()
    y2 = random()
    return ( (x1 - x2) ** 2 + (y1 - y2) ** 2 ) ** 0.5

f = 0
N = 10000
for k in range(1000):
    m = 0
    for k in range(N):
        m += dist()
    f += m/N
print(f / 1000)

```

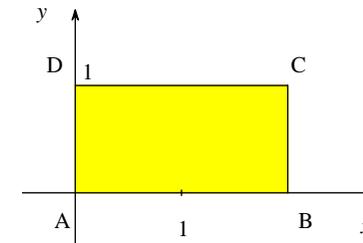
“0.5213943752132129” la première fois
 “0.5208317781716778” une deuxième fois
 “0.5213418328109504” une troisième fois

Je vais m'arrêter là car je vois que les trois valeurs ne sont pas éloignées du tout. Je peux donc raisonnablement conclure qu'en moyenne, les deux points sont à une distance d'à peu près 0,521 l'un de l'autre.

Avec la calculatrice, $\frac{2 + \sqrt{2} + 5 \ln(\sqrt{2} + 1)}{15} = 0,521405433\dots$

3°) distance moyenne dans un cube d'arête 1 : 0,6620196846610... (valeur obtenue le 30-5-2022)

Cas d'un rectangle de longueur 2 et de largeur 1 :



Programme Vincente Seixas le 30-5-2022 :

```

from math import sqrt
from random import uniform

def distance2d(xa, xb, ya, yb):
    d=sqrt((xa-xb)**2+(ya-yb)**2)
    return d

d=0
for i in range(1000):
    r=distance2d(uniform(0, 2), uniform(0, 1), uniform(0, 1), uniform(0, 1))
    d=d+r
print('moyenne: ', d/1000)

```

On trouve des valeurs autour de 0,79 – 0,81.

```

from random import uniform

def lancer(n, p):
    t=0
    compt=0
    tour=0
    for i in range(n):
        while compt!=2:
            res=uniform(0, 1)
            if res<=p:
                compt+=1
                tour+=1
            else:
                compt=0
                tour+=1

    compt=0
    t+=1
    return tour/n

```

Exercice **22**: Voir démonstration DS 2 15 octobre 2022 ECG 2^e année Anthony Mansuy

22

Le 30-5-2022

Proposition pour une probabilité du type $1/p$ (Vicente Seixas)

```

from random import random

def exo22(n, p):
    L=[]
    for i in range(n):
        x=0
        a=0
        b=randint(1, p)
        while a!=1 or b!=1:
            a=b
            b=randint(1, p)
            x+=1
        L.append(x)
    return sum(L)/len(L)

```

22

23

Pour la valeur absolue, on fait appel à la fonction abs présente dans Python (il n'y a pas à l'importer).

```

from random import uniform

def dist():
    d=abs(uniform(0, 1)-uniform(0, 1))
    return d

```

```

def estim_dist_moy(n):
    L=[dist() for _ in range(n)]
    return sum(L)/n

```

On trouve une valeur proche de 0,33.

On peut démontrer que la valeur théorique est $\frac{1}{3}$.

Ancien programme (un peu moins bien) :

```
from random import uniform

def estimdist(n):
    d=0
    for i in range(n) :
        x,y= uniform(0,1), uniform(0,1)
        r=abs(x-y)
        d=d+r
    return d/n
```

```
from random import random

def simuY(p) :
    n=0
    while random()<p:
        n=n+1
    while random()<p:
        n=n+1
    return n
```

Autre proposition d'après EML 2018 corrigé Major-Prepa

```
from random import random

def simuY(p) :
    x=0
    n=0
    while n<2:
        if random()<p:
            n=n+1
        else :
            x=x+1
    return x
```

24

Premier jeu :

```
from random import random

def simuX(p) :
    n=0
    while random()<p:
        n=n+1
    return n
```

Deuxième jeu :

Source exercice 24 : Jérémy Legendre Concours blanc 14 mai 2022 exercice 2

Deux joueurs, A et B, décident de s'affronter dans un jeu de hasard.

- Le joueur A lance une pièce donnant PILE avec la probabilité p .
- Sujets enregistrés le 11-7-2022 d'après sujets Ecricome 2014 voie économique

25

1°) a) Donner les valeurs de $P(X_i = 0)$ et de $P(X_i = 1)$.

On peut faire un tableau donnant la loi de probabilité.

X_i suit la loi binomiale de paramètre p .

b) Justifier que : $P(Y_i = 1) = P(X_{i-1} = 0) \times P(X_i = 1) + P(X_{i-1} = 1) \times P(X_i = 0)$.

c) Déterminer $P(Y_i = 1)$ en fonction de p .

$$P(Y_i = 1) = (1-p) \times p + p \times (1-p)$$

$$P(Y_i = 1) = 2p(1-p)$$

3°)

```
from random import randint

def blocs01():
    L=[rand(0, 1) for k in range(1001)]
    s=(L[1]!=L[0])
    for i in range(1, 1000):
        s=s+(L[i]!=L[i+1])
    return (1+s)
```

25

```
def blocs01(L):
    t=(L[1] !=L[0])
    for i in range(1, len(L)):
        t=t+(L[i] !=L[i-1])
    return t
```

blocs01([0,1,1,0,1,1,1])

4

blocs01([0,1,1])

2

blocs01([0,1,0,1])

4

blocs01([0,1,1,1,1])

2

Deuxième fonction : tirage d'un nombre 0 ou 1

```
from random import random

def tirage():
    r=random
    if random<p:
        x=1
    else:
        x=0
    return x
```

Programme final :

```
from random import randint

def simuS(N):
    L=[tirage() for k in range(N)]
    s=blocs01(N)
    return s
```

- Recopier ce programme et compléter la ligne `L=[... for k in range(1001)]`.
- Quel est le rôle de la ligne `s=(L[1] !=L[0])` ?
- Compléter la ligne `S=S+...` à l'aide d'un booléen.

Il s'agit du « nombre moyen de séries » dans une suite de pile ou face.

Le 3 juin 2021

J'avais noté TS « Les méthodes de Monte-Carlo ». J'en avais en fait déjà parlé durant l'année.

Fin des solutions des exercices

Source :

10 100 personnes jouent indépendamment à un même jeu dont la variable aléatoire

Exercice : (livre Sésamath Terminale spécialité)

Exercice fondamental : livre terminale Sésamath spécialité

La probabilité que la moyenne s'éloigne de l'espérance devient très petite pour les grands échantillons.

Pt1urcpel'InégalitéA B snil viiifiée sactnnt que
A >2 B. Il sulfit d'av0'r
B 0.95.

Le 6 avril 2021

Fabien Herbaut Feuille 4 sur Bienaymé-Tchebycheff

Le 6 mai 2021

Bienaymé-Tchebycheff
Cours Johanne Cohen.

Markov

Extrait du programme :

L'objectif de cette section est d'une part d'approfondir le sens de l'écart-type comme mesure de dispersion, d'autre part de couronner la partie « Probabilités » par la loi des grands nombres, qui est le premier résultat fondamental de la théorie des probabilités et dont les implications sont considérables. Pour cela, l'outil employé est l'inégalité de Bienaymé-Tchebychev dont l'idée fondamentale est mise en valeur : l'écart type d'une variable aléatoire X est l'unité naturelle pour étudier la dispersion de X autour de son espérance

- ; par construction, il est naturel d'observer des écarts de X à
- en deçà ou au-delà de □ .

L'inégalité de Bienaymé-Tchebychev montre qu'en revanche des écarts de X à □ de quelques □ deviennent improbables. Ce résultat, d'une importance majeure en lui-même, permet de plus d'établir la loi des grands nombres, selon laquelle l'écart entre la moyenne d'un échantillon d'une variable aléatoire et l'espérance de cette variable ne dépasse une valeur donnée à l'avance qu'avec une probabilité qui tend vers zéro quand la taille de l'échantillon tend vers l'infini.

Il est utile de faire remarquer aux élèves que le caractère universel de l'inégalité de Bienaymé-Tchebychev a pour contrepartie le fait qu'elle est loin d'être optimale : ainsi, elle montre qu'un écart à μ supérieur à 2σ est de probabilité inférieure ou égale à $\frac{1}{4}$ alors que les élèves ont découvert par simulation que cette probabilité est souvent majorée par 0,05.

En avoir conscience ne diminue pas l'intérêt théorique de l'inégalité de Bienaymé-Tchebychev, et permet de mettre en évidence des cas de raisonnement par conditions suffisantes, par exemple la recherche d'une taille d'échantillon pour majorer une probabilité.

Loi des grands nombres

Inégalité de Bienaymé-Tchebychev

La principale interprétation de l'écart-type σ réside dans le fait que les écarts avec l'espérance sont de l'ordre de grandeur de σ . L'inégalité de Bienaymé-Tchebychev précise cette idée.

INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV Soit X une variable aléatoire d'espérance μ et de variance V . Pour tout réel strictement positif δ , on a : $P(|X - \mu| \geq \delta) \leq \frac{V(X)}{\delta^2}$.

DÉMONSTRATION

Soit $\Omega' = \{x_1; x_2; \dots; x_n\}$ l'ensemble des valeurs prises par la variable aléatoire X et soit l'événement $A = \{|X - \mu| > \delta\}$, dont on note les éléments a_1, a_2, \dots, a_k .

Ces éléments a_1, a_2, \dots, a_k appartiennent à A , et $|a_i - \mu| > \delta$ pour tout i ($1 \leq i \leq k$).

On sait que, $V(X) = (x_1 - \mu)^2 P(X = x_1) + \dots + (x_n - \mu)^2 P(X = x_n)$.

Tous les termes de cette somme sont positifs, donc si on ne garde dans cette somme que les termes portant sur les issues présentes dans A , la somme diminue, d'où :

$$V(X) \geq (a_1 - \mu)^2 P(X = a_1) + \dots + (a_k - \mu)^2 P(X = a_k).$$

Par définition de A , chaque nombre $|a_i - \mu|$ est supérieur à δ , donc chacun des nombres $(a_i - \mu)^2$ est supérieur à δ^2 . Ainsi :

$$(a_1 - \mu)^2 P(X = a_1) + \dots + (a_k - \mu)^2 P(X = a_k) \geq \delta^2 P(X = a_1) + \dots + \delta^2 P(X = a_k).$$

$$V(X) \geq \delta^2 (P(X = a_1) + \dots + P(X = a_k)) \text{ et } P(X = a_1) + \dots + P(X = a_k) = P(X \in A).$$

D'où : $V(X) \geq \delta^2 P(X \in A)$, et ainsi $P(X \in A) \leq \frac{V(X)}{\delta^2}$, soit encore $P(|X - \mu| \geq \delta) \leq \frac{V(X)}{\delta^2}$.

INÉGALITÉ DE CONCENTRATION Soit M_n la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V .

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

INÉGALITÉ DE CONCENTRATION Soit M_n la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V .

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

DÉMONSTRATION

On applique l'inégalité de Bienaymé-Tchebychev à M_n :

$$P(|M_n - E(M_n)| \geq \delta) \leq \frac{V(M_n)}{\delta^2}. \text{ D'une part, } E(M_n) = \mu \text{ et d'autre part } V(M_n) = \frac{V}{n},$$

ce qui entraîne $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

Loi des grands nombres

Inégalité de Bienaymé-Tchebychev

La principale interprétation de l'écart-type σ réside dans le fait que les écarts avec l'espérance sont de l'ordre de grandeur de σ . L'inégalité de Bienaymé-Tchebychev précise cette idée.

INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV Soit X une variable aléatoire d'espérance μ et de variance V . Pour tout réel strictement positif δ , on a : $P(|X - \mu| \geq \delta) \leq \frac{V(X)}{\delta^2}$.

DÉMONSTRATION

Soit $\Omega' = \{x_1; x_2; \dots; x_n\}$ l'ensemble des valeurs prises par la variable aléatoire X et soit l'événement $A = \{|X - \mu| \geq \delta\}$, dont on note les éléments a_1, a_2, \dots, a_k .

Ces éléments a_1, a_2, \dots, a_k appartiennent à A , et $|a_i - \mu| \geq \delta$ pour tout i ($1 \leq i \leq k$).

On sait que, $V(X) = (x_1 - \mu)^2 P(X = x_1) + \dots + (x_n - \mu)^2 P(X = x_n)$.

Tous les termes de cette somme sont positifs, donc si on ne garde dans cette somme que les termes portant sur les issues présentes dans A , la somme diminue, d'où :

$$V(X) \geq (a_1 - \mu)^2 P(X = a_1) + \dots + (a_k - \mu)^2 P(X = a_k).$$

Par définition de A , chaque nombre $|a_i - \mu|$ est supérieur à δ , donc chacun des nombres $(a_i - \mu)^2$ est supérieur à δ^2 . Ainsi :

$$(a_1 - \mu)^2 P(X = a_1) + \dots + (a_k - \mu)^2 P(X = a_k) \geq \delta^2 P(X = a_1) + \dots + \delta^2 P(X = a_k).$$

$V(X) \geq \delta^2 (P(X = a_1) + \dots + P(X = a_k))$ et $P(X = a_1) + \dots + P(X = a_k) = P(X \in A)$.

D'où : $V(X) \geq \delta^2 P(X \in A)$, et ainsi $P(X \in A) \leq \frac{V(X)}{\delta^2}$, soit encore $P(|X - \mu| \geq \delta) \leq \frac{V(X)}{\delta^2}$.

INÉGALITÉ DE CONCENTRATION Soit M_n la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V .

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

DÉMONSTRATION

On applique l'inégalité de Bienaymé-Tchebychev à M_n :

$$P(|M_n - E(M_n)| \geq \delta) \leq \frac{V(M_n)}{\delta^2}. \text{ D'une part, } E(M_n) = \mu \text{ et d'autre part } V(M_n) = \frac{V}{n},$$

ce qui entraîne $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

Loi faible des grands nombres

THÉORÈME Soit M_n la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V .

Alors pour tout réel $\delta > 0$, on a : $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.

DÉMONSTRATION

Soit δ un réel strictement positif. D'après l'inégalité de concentration, pour tout entier naturel n non nul, $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

D'où : $0 \leq P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$ et $\lim_{n \rightarrow +\infty} \left(\frac{V}{n\delta^2}\right) = 0$.

Avec le théorème des gendarmes, on en déduit : $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.

Ce résultat justifie la possibilité de définir des probabilités en prenant pour valeurs approchées les fréquences obtenues pour un « grand » nombre d'essais.

Loi faible des grands nombres

THÉORÈME Soit M_n la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V .

Alors pour tout réel $\delta > 0$, on a : $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.

DÉMONSTRATION

Soit δ un réel strictement positif. D'après l'inégalité de concentration, pour tout entier naturel n non nul, $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$.

D'où : $0 \leq P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$ et $\lim_{n \rightarrow +\infty} \left(\frac{V}{n\delta^2}\right) = 0$.

Avec le théorème des gendarmes, on en déduit : $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$.

Ce résultat justifie la possibilité de définir des probabilités en prenant pour valeurs approchées les fréquences obtenues pour un « grand » nombre d'essais.

Point Histoire

Irénée-Jules Bienaymé (1796-1878) est un mathématicien français qui a énoncé cette inégalité en 1853. Mais, c'est Pafnouti Tchebychev (1821-1894) qui l'a démontrée en 1867.

Point Histoire



C'est Jacques Bernoulli (1654-1705) 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.

À noter

La probabilité que la moyenne s'éloigne de l'espérance devient très petite pour les grands échantillons.

Vidéo 08:04

Utiliser l'inégalité de Bienaymé-Tchebychev



6 CAPACITÉ

Utiliser l'inégalité de concentration pour définir une taille d'échantillon

On effectue n tirages successifs, avec remise, d'une boule dans une urne contenant 2 boules rouges et 3 boules noires. On note X la variable aléatoire qui, à un tirage donné, associe 1 si la boule tirée est rouge, et 0 sinon, et M_n la variable aléatoire moyenne d'un échantillon de taille n de X .

- Déterminer $E(X)$ et $V(X)$, puis écrire l'inégalité de concentration relative à M_n .
- À partir de quel nombre de tirages peut-on garantir à plus de 95 % que la proportion de boules rouges obtenues restera strictement comprise entre 0,35 et 0,45 ?

Méthode

Pour que l'inégalité $A \leq B$ soit vérifiée sachant que $A \leq B$, il suffit d'avoir $B \leq 0,95$.

LOGIQUE

On utilise ici un raisonnement par condition suffisante, voir p. 449.

Voir exercices

- 80 Pour s'entraîner
- 98 Pour travailler en autonomie

Solution commentée

1 X est une variable aléatoire de Bernoulli de paramètre $\frac{2}{5} = 0,4$ donc $E(X) = 0,4$ et $V(X) = 0,4 \times (1 - 0,4) = 0,24$. On écrit l'inégalité de concentration relative à M_n , variable aléatoire moyenne d'un échantillon de taille n de X avec $\mu = 0,4$ et $V = 0,24$:

$$P(|M_n - 0,4| \geq \delta) \leq \frac{0,24}{n\delta^2}, \text{ pour tout réel } \delta > 0.$$

2 On cherche ici le plus petit entier n tel que $P(0,35 < M_n < 0,45) > 0,95$.

On remarque que $0,35 < M_n < 0,45$ équivaut à $|M_n - 0,4| < 0,05$.

On veut que M_n vérifie : $P(|M_n - 0,4| < 0,05) > 0,95$.

L'événement contraire de $\{|M_n - 0,4| < 0,05\}$ est $\{|M_n - 0,4| \geq 0,05\}$.

Ainsi : $P(|M_n - 0,4| < 0,05) = 1 - P(|M_n - 0,4| \geq 0,05)$.

Donc : $P(|M_n - 0,4| < 0,05) > 0,95$ équivaut à $1 - P(|M_n - 0,4| \geq 0,05) > 0,95$ soit $P(|M_n - 0,4| \geq 0,05) < 0,05$ (1).

L'inégalité de concentration avec $\delta = 0,05$ s'écrit :

$$P(|M_n - 0,4| \geq 0,05) \leq \frac{0,24}{n \times 0,05^2}, \text{ soit } P(|M_n - 0,4| \geq 0,05) \leq \frac{60\,000}{n}.$$

L'inégalité (1) sera vérifiée si l'on choisit n tel que $\frac{60\,000}{n} < 0,05$.

On en déduit : $0,05n > 60\,000$, soit $n > \frac{60\,000}{0,05}$.

$\frac{60\,000}{0,05} = 1\,200\,000$, donc le nombre minimal de lancers est 1 200 001.

Inégalité et interprétation de Bienaymé-Tchebycheff

A Inégalité et interprétation

Propriété

X est 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 : • Pour une démonstration, voir exercice 23 page 496.

• Pour $\delta \leq \sigma$ où σ est l'écart-type de X , $\frac{V}{\delta^2} \geq 1$ et l'inégalité est évidente.

Exemple et interprétation

• On lance une pièce équilibrée 100 fois de suite, la variable aléatoire X donne le nombre de Pile obtenus.

X suit la loi binomiale de paramètres $n = 100$ et $p = \frac{1}{2}$.

l'espérance de X est $\mu = 100 \times \frac{1}{2} = 50$ et sa variance est $V = 100 \times \frac{1}{2} \times \frac{1}{2} = 25$.

D'après l'inégalité de Bienaymé-Tchebychev, pour tout réel $\delta > 0$,

$$P(|X - 50| \geq \delta) \leq \frac{25}{\delta^2}.$$

• Par exemple pour $\delta = 10$, $P(|X - 50| \geq 10) \leq 0,25$.

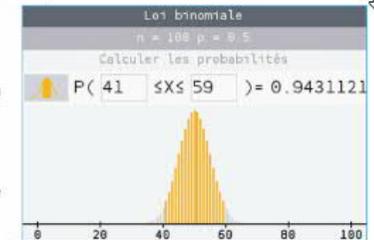
La probabilité que l'écart de X à $\mu = 50$ soit supérieur ou égal à 10 est inférieure à 0,25.

Or, $P(|X - 50| < 10) = 1 - P(|X - 50| \geq 10)$,

donc $P(|X - 50| < 10) \geq 0,75$.

La probabilité que X prenne une valeur dans l'intervalle $]40 ; 60[$ est supérieure à 0,75.

Vidéo
JAI COMPRIS.COM
Cette notion est présentée en vidéo



B Des situations particulières

X est une variable aléatoire d'espérance μ et d'écart-type σ .

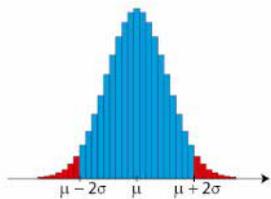
On applique l'inégalité de Bienaymé-Tchebychev avec $\delta = k\sigma$ où k est un nombre entier naturel, $k \geq 2$:

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

Par exemple :

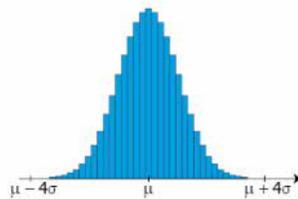
• $k = 2, \quad P(|X - \mu| \geq 2\sigma) \leq 0,25$

La probabilité que l'écart de X à μ soit supérieur ou égal à 2σ est inférieure ou égale à 0,25.



• $k = 4, \quad P(|X - \mu| \geq 4\sigma) \leq 0,0625$

Obtenir un écart de X à μ supérieur ou égal à 4σ est un événement improbable.



1 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.

1. a) Donner la loi de probabilité de la variable aléatoire X.
- b) Déterminer l'espérance μ et l'écart-type σ de X.
- c) Établir à l'aide de l'inégalité de Bienaymé-Tchebychev que :

$$P(|X - \mu| \geq 2\sigma) \leq 0,25.$$

2. On écrit un programme en langage Python afin de simuler un échantillon de taille n de la variable aléatoire X.

Interpréter le résultat renvoyé par :

- a) la fonction **Distance** ; b) la fonction **Echantillon**.
3. a) Saisir ce programme et exécuter **Echantillon(1000)**.
- b) Comparer au résultat obtenu avec l'inégalité de Bienaymé-Tchebychev.

```

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
    
```

Solution

tion

1. a) La variable aléatoire X suit la loi binomiale de paramètres 100 et 0,7.

b) $\mu = 100 \times 0,7 = 70$ et $\sigma = \sqrt{100 \times 0,7 \times 0,3} = \sqrt{21}$ soit $\sigma = 4,58$.

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

Avec $\delta = 2\sigma$, on obtient $P(|X - \mu| \geq 2\sigma) \leq 0,25$.

2. a) Lors d'une simulation de X, la fonction **Distance** renvoie l'écart $|x - 70|$ où x est la valeur prise par la variable aléatoire X.

b) Pour un échantillon de taille n de X, la fonction **Echantillon** renvoie la proportion des valeurs x prises par X telles que $|x - 70| \geq 2\sigma$.

3. a) Par exemple, on obtient :

```

>>> Echantillon(1000)
0.038
>>> Echantillon(1000)
0.043
>>> Echantillon(1000)
0.048
    
```

b) On observe que dans moins de 5 % des cas $|x - 70| \geq 2\sigma$.

L'inégalité de Bienaymé-Tchebychev donne donc un résultat qui n'est pas optimal.

Inégalité de Bienaymé-Tchebychev

Cours 1

Questions Flash À l'oral

9 X est une variable aléatoire d'espérance $\mu = 50$.

Pour tout réel δ , laquelle de ces probabilités l'inégalité de Bienaymé-Tchebychev permet-elle de majorer ?

- (1) $P(|X - 50| = \delta)$ (2) $P(|X - 50| \geq \delta)$
 (3) $P(|X - 50| \leq \delta)$ (4) $P(|X| \leq 50)$

10 X est une variable aléatoire d'espérance μ et d'écart-type σ .

Laquelle de ces inégalités obtient-on avec l'inégalité de Bienaymé-Tchebychev ?

- (1) $P(|X - \mu| \geq 2\sigma) \leq 0,05$
 (2) $P(|X - \mu| \geq 2\sigma) \leq 0,25$
 (3) $P(|X - \mu| \leq 2\sigma) \leq 0,25$

Inégalité de Bienaymé-Tchebychev

Cours 1

Questions Flash À l'oral

9 X est une variable aléatoire d'espérance $\mu = 50$.

Pour tout réel δ , laquelle de ces probabilités l'inégalité de Bienaymé-Tchebychev permet-elle de majorer ?

- (1) $P(|X - 50| = \delta)$ (2) $P(|X - 50| \geq \delta)$
 (3) $P(|X - 50| \leq \delta)$ (4) $P(|X| \leq 50)$

10 X est une variable aléatoire d'espérance μ et d'écart-type σ .

Laquelle de ces inégalités obtient-on avec l'inégalité de Bienaymé-Tchebychev ?

- (1) $P(|X - \mu| \geq 2\sigma) \leq 0,05$
 (2) $P(|X - \mu| \geq 2\sigma) \leq 0,25$
 (3) $P(|X - \mu| \leq 2\sigma) \leq 0,25$

11 Une variable aléatoire X a pour espérance $\mu = 250$ et pour variance $V = 125$.

a) À l'aide de l'inégalité de Bienaymé-Tchebychev, établir que :

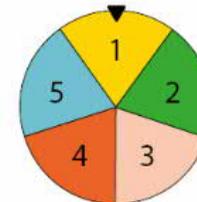
$$P(|X - 250| \geq 20) \leq 0,3125.$$

b) En déduire alors que :

$$P(|X - 250| < 20) \geq 0,6875.$$

c) Interpréter cette dernière inégalité.

12 La roue équilibrée ci-contre est partagée en cinq secteurs identiques numérotés de 1 à 5.



On fait tourner la roue 100 fois de suite, X est la variable aléatoire qui donne le nombre de fois où le 1 est sorti.

- a) Donner la loi de probabilité de X.
 b) Déterminer l'espérance et la variance de X.
 c) Justifier que pour tout réel $\delta > 0$,

$$P(|X - 20| \geq \delta) \leq \frac{16}{\delta^2}.$$

d) En déduire que la probabilité de l'événement « X prend une valeur en dehors de l'intervalle [11 ; 29] » est inférieure ou égale à 0,16.

JAI COMPRIS.COM

Cet exercice est corrigé en vidéo



13 X est une variable aléatoire d'espérance μ et d'écart-type σ .

1. Justifier que pour tout entier naturel $k \geq 1$,

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

2. a) Déterminer une valeur de k pour laquelle $P(|X - \mu| \geq k\sigma) \leq 0,1$.

b) Interpréter la valeur de k obtenue.

Un service de livraison constate que 5 % des colis livrés sont abimés.

Pendant sa journée de travail, Louis livre 100 colis. On note X le nombre de colis abimés parmi ces 100 colis.

1. a) Donner la loi de probabilité de X.
 b) Déterminer l'espérance μ et l'écart-type σ de X.
 c) Établir que :

$$P(|X - 5| \geq 2\sigma) \leq 0,25.$$

2. Avec le programme suivant, on réalise une simulation d'un échantillon de taille n de la variable aléatoire X.

```

1 from random import *
2 from math import *
3
4 def Ecart():
5     x=0
6     for k in range(100):
7         a=random()
8         if a<=0.05:
9             x=x+1
10        d=abs(x-5)
11        return d
12
13 def Echantillon(n):
14     s=sqrt(4.75)
15     y=0
16     for j in range(n):
17         if Ecart()>=2*s:
18             y=y+1
19     p=y/n
20     return p

```

Interpréter le résultat renvoyé par la fonction

Interpréter le résultat renvoyé par la fonction **Echantillon**.

3. a) Saisir le programme.

b) Exécuter plusieurs fois **Echantillon(1000)**. Commenter les résultats obtenus.

4. Comparer les résultats obtenus par simulation avec celui donné par l'inégalité de Bienaymé-Tchebychev.

23 Démontrer l'inégalité de Bienaymé-Tchebychev

On a admis en cours au paragraphe 1. A la propriété suivante :

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

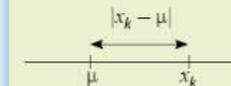
$$\text{Pour tout réel } \delta > 0, P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}.$$

Rédiger la démonstration de cette propriété en suivant le guide ci-dessous.

On donne un nombre réel $\delta > 0$, on note x_1, x_2, \dots, x_n les valeurs prises par la variable aléatoire X et on désigne par H l'ensemble des valeurs x_k telles que $|x_k - \mu| \geq \delta$.

- (1) Écrire la définition de la variance de X : $V = \sum_{i=1}^n P(X = x_i)(\dots)^2$
- (2) Minorer V : $V \geq \sum_{x_k \in H} P(X = x_k)(\dots)^2$
- (3) Tirer une conséquence : $V \geq \dots^2 \sum_{x_k \in H} P(X = x_k)$
- (4) Compléter : Or, $\sum_{x_k \in H} P(X = x_k) = \dots$ donc $V \geq \dots$
- (5) Conclure : $P(H) \leq \dots$. Rédiger une phrase de conclusion.

Conseil



$x_k \in H$ si, et seulement si,
 $|x_k - \mu| \geq \delta$.

Conseil

$\sum_{x_k \in H} \dots$ signifie
qu'on effectue
la somme des quantités
indiquées pour chacune
des valeurs x_k de H.

24 Estimer une probabilité

p est la probabilité inconnue du succès d'une épreuve de Bernoulli. On répète 4 040 fois cette épreuve et on obtient 2 049 succès. Voici le raisonnement qui permet d'estimer p avec une probabilité au moins égale à 0,95.

Lorsqu'on répète 4 040 fois l'épreuve, on note F la fréquence des succès.

L'espérance de la variable aléatoire F est égale à p .

La variance de la variable aléatoire F est égale à $\frac{p(1-p)}{4\,040}$.

D'après l'inégalité de Bienaymé-Tchebychev, pour tout réel $\delta > 0$,

$$P(|F - p| < \delta) \geq 1 - \frac{p(1-p)}{4\,040\delta^2}$$

Le maximum de la fonction $f: x \mapsto x(1-x)$ sur l'intervalle $[0; 1]$ est $\frac{1}{4}$.

On en déduit alors que pour tout réel $\delta > 0$,

$$P(|F - p| < \delta) \geq 1 - \frac{1}{16\,160\delta^2}$$

Puis, on en déduit que $P(|F - p| < 0,05) \geq 0,95$.

$p \approx \frac{2\,049}{4\,040}$ à 0,05 près avec une probabilité supérieure à 0,95.

27 On écrit au hasard une suite de n chiffres.

X est la variable aléatoire qui donne le nombre de fois où le chiffre 5 apparaît.

1. a) Donner la loi de probabilité de X .

b) Déterminer l'espérance et la variance de X .

2. F est la variable aléatoire qui donne la fréquence du chiffre 5 dans l'écriture.

a) Justifier que $F = \frac{X}{n}$.

b) Déterminer l'espérance de la variable aléatoire F .

c) Démontrer que la variance de F est $\frac{0,09}{n}$.

d) Que dire de la variance de F lorsque n prend des valeurs de plus en plus grandes.

3. a) Écrire l'inégalité de Bienaymé-Tchebychev pour la variable aléatoire F .

b) En déduire que :

$$P(0,05 < F < 0,15) \geq 1 - \frac{36}{n}$$

d) Que dire de la variance de F lorsque n prend des valeurs de plus en plus grandes.

3. a) Écrire l'inégalité de Bienaymé-Tchebychev pour la variable aléatoire F .

b) En déduire que :

$$P(0,05 < F < 0,15) \geq 1 - \frac{36}{n}$$

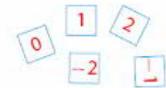
c) Déterminer n tel que la fréquence d'apparition du chiffre 5 soit comprise entre 5 % et 15 % avec une probabilité supérieure à 0,9.

31 SIMULATION DE n ÉCHANTILLONS DE TAILLE n D'UNE VARIABLE ALÉATOIRE.

Algo python

CALCUL DE L'ÉCART-TYPE DE LA SÉRIE DES MOYENNES DES ÉCHANTILLONS

Dans une enveloppe, on place cinq cartons indiscernables portant les numéros $-2, -1, 0, 1$ et 2 . On tire au hasard un carton et on relève son numéro. La variable aléatoire X donne pour valeur le carré du numéro tiré.



1. Loi de probabilité de X

a) Déterminer la loi de probabilité de X .

b) Calculer l'espérance μ et l'écart-type σ de X .

c) Quel résultat la fonction X écrite en langage Python renvoie-t-elle ?

2. Simulation d'un échantillon de taille n de X

La fonction `Moyenne` a pour paramètre la taille n d'un échantillon de la variable aléatoire X .

Interpréter le résultat renvoyé par cette fonction.

```
1 from random import *
2 from math import *
3
4 def X():
5     a=randint(-2,2)
6     x=a**2
7     return x
8
9 def Moyenne(n):
10    somme=0
11    for i in range(n):
12        somme=somme+X()
13    m=somme/n
14    return m
15
```

46 L'inégalité de Markov

X est une variable aléatoire dont les valeurs x_1, x_2, \dots, x_n sont positives.

α désigne un nombre réel strictement positif.

a) Justifier l'écriture :

$$E(X) = \sum_{x_j \geq \alpha} P(X = x_j)x_j + \sum_{x_j < \alpha} P(X = x_j)x_j.$$

b) En déduire que $E(X) \geq \sum_{x_j \geq \alpha} x_j P(X = x_j)$.

c) Établir l'inégalité de Markov :

$$P(X \geq \alpha) \leq \frac{E(X)}{\alpha}.$$

48 Une proportion inconnue

Une usine fabrique des roulettes en plastique dont une proportion inconnue p est défectueuse ; on souhaite trouver une valeur approchée de p .



On effectue un prélèvement de n roulettes, on assimile ce prélèvement à n tirages indépendants avec remise. On note X_n la variable aléatoire égale au nombre de pièces défectueuses dans le lot.

1. a) Donner la loi de probabilité de X_n .

b) Déterminer son espérance et sa variance.

2. a) Démontrer que pour tout réel $\varepsilon > 0$,

1. a) Donner la loi de probabilité de X_n .

b) Déterminer son espérance et sa variance.

2. a) Démontrer que pour tout réel $\varepsilon > 0$,

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

b) Étudier le sens de variation de la fonction f définie sur l'intervalle $[0; 1]$ par $f(x) = x(1-x)$.

En déduire que $p(1-p) \leq \frac{1}{4}$.

c) Démontrer alors que pour tout réel $\varepsilon > 0$,

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

3. En déduire une condition sur n pour qu'une valeur effective de $\frac{X_n}{n}$ soit une valeur approchée de p à 10^{-2} près avec une probabilité supérieure à 95 %.

Livre Hyperbole édition 2020

La loi géométrique

4. Loi géométrique

Définition, espérance

DÉFINITION On considère une épreuve de Bernoulli de paramètre p dont le succès est noté S. Soit X la variable aléatoire comptant le nombre de répétitions nécessaires de cette épreuve de Bernoulli avant l'obtention d'un premier succès S. Cette variable aléatoire prend des valeurs entières non nulles.

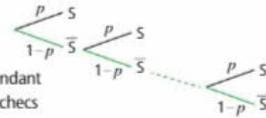
La loi de probabilité de X est appelée **loi géométrique** de paramètre p .

EXEMPLE: On lance une pièce de monnaie jusqu'à la première obtention du côté Pile. La variable aléatoire X , comptant le nombre de lancers nécessaires pour l'obtention du premier Pile, suit la loi géométrique de paramètre 0,5.

PROPRIÉTÉ Soit X une variable aléatoire suivant la loi géométrique de paramètre p . Pour tout entier naturel k non nul, $P(X = k) = (1 - p)^{k-1} \times p$.

DÉMONSTRATION

On souhaite obtenir le premier succès après k répétitions. On est en présence d'un schéma de Bernoulli de paramètres k et p . Dans l'arbre correspondant à ce schéma, un seul chemin est constitué de $k - 1$ échecs suivi d'un succès : sa probabilité vaut $(1 - p)^{k-1} \times p$.

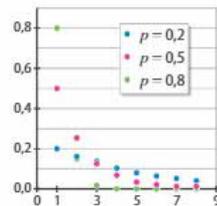


PROPRIÉTÉ Soit X une variable aléatoire suivant la loi géométrique de paramètre p . Alors on a $E(X) = \frac{1}{p}$.

PROPRIÉTÉ Soit X une variable aléatoire suivant la loi géométrique de paramètre p . Alors on a $E(X) = \frac{1}{p}$.

Représentation graphique

Voici les représentations graphiques de trois lois géométriques de paramètres 0,2, 0,5 et 0,8. On remarque que plus le nombre k est grand, plus la probabilité $P(X = k)$ est faible.



Propriétés caractéristiques

PROPRIÉTÉ Soit X une variable aléatoire.

- (1) Si X suit une loi géométrique, alors X est sans mémoire, c'est-à-dire que pour tous entiers m et n non nuls, on a $P_{X>n}(X > m + n) = P(X > m)$.
- (2) Réciproquement, si X est sans mémoire, alors X suit une loi géométrique.

DÉMONSTRATION

(1) Soit X une variable aléatoire suivant la loi géométrique de paramètre p . $P(X > m) = 1 - P(X \leq m) = 1 - [P(X = 1) + \dots + P(X = m)]$.

$$\text{Or } P(X = 1) + \dots + P(X = m) = p + (1-p)p + \dots + (1-p)^{m-1} \times p$$

$$= p \times [1 + (1-p) + \dots + (1-p)^{m-1}] = p \times \frac{1 - (1-p)^m}{1 - (1-p)} = 1 - (1-p)^m.$$

Car $1 + (1-p) + \dots + (1-p)^{m-1}$ est la somme des m premiers termes d'une suite géométrique. Ainsi $P(X > m) = 1 - (1 - (1-p)^m) = (1-p)^m$.

$$\text{On a } P_{X>n}(X > m + n) = \frac{P(\{X > m + n\} \cap \{X > n\})}{P(X > n)} = \frac{P(X > m + n)}{P(X > n)}$$

$$= \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X > m).$$

DÉMONSTRATION

(1) Soit X une variable aléatoire suivant la loi géométrique de paramètre p . $P(X > m) = 1 - P(X \leq m) = 1 - [P(X = 1) + \dots + P(X = m)]$.

$$\text{Or } P(X = 1) + \dots + P(X = m) = p + (1-p)p + \dots + (1-p)^{m-1} \times p$$

$$= p \times [1 + (1-p) + \dots + (1-p)^{m-1}] = p \times \frac{1 - (1-p)^m}{1 - (1-p)} = 1 - (1-p)^m.$$

Car $1 + (1-p) + \dots + (1-p)^{m-1}$ est la somme des m premiers termes d'une suite géométrique. Ainsi $P(X > m) = 1 - (1 - (1-p)^m) = (1-p)^m$.

$$\text{On a } P_{X>n}(X > m + n) = \frac{P(\{X > m + n\} \cap \{X > n\})}{P(X > n)} = \frac{P(X > m + n)}{P(X > n)}$$

$$= \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X > m).$$

À noter

La variable aléatoire X représente le « temps d'attente » de la réalisation du premier succès S.

À noter

Ici, on a démontré un lemme, c'est-à-dire un résultat intermédiaire sur lequel on s'appuie pour conduire la démonstration.

Identifier une situation où une variable aléatoire suit une loi géométrique

Dans chacun des cas suivants, déterminer si la variable aléatoire X proposée suit ou non une loi géométrique. Si oui, préciser alors le paramètre de cette loi.

- 1 Une urne contient une boule blanche et deux noires. On tire, successivement et avec remise, n boules de cette urne. On note X la variable aléatoire comptant le nombre de boules blanches obtenues.
- 2 On lance un dé équilibré à six faces plusieurs fois de suite : tant qu'on n'a pas obtenu un 6, on continue à lancer le dé. X est la variable aléatoire qui compte le nombre de répétitions nécessaires à l'obtention de la valeur 6.
- 3 Sur une route départementale fréquentée, on estime qu'un autopisteur a 15 % de chances d'être pris à chaque fois qu'un véhicule passe à côté de lui. X est le nombre de véhicules qu'il a vu au total.

Solution commentée

- 1 X compte le nombre de succès obtenus lors d'une série de n tirages. On peut très bien ne tirer aucune boule blanche, donc X ne prend pas ses valeurs dans l'ensemble des entiers non nuls. X ne suit pas une loi géométrique mais la loi binomiale de paramètres n et $\frac{1}{3}$.
- 2 La variable aléatoire X compte le nombre de répétitions nécessaires de l'expérience jusqu'à l'obtention du succès « obtenir un 6 ». Elle prend ses valeurs dans l'ensemble des entiers non nuls. La probabilité de succès vaut $\frac{1}{6}$.
 X suit bien la loi géométrique de paramètre $\frac{1}{6}$.
- 3 La variable aléatoire X compte le nombre de répétitions nécessaires de l'expérience jusqu'à l'obtention du succès « l'autostoppeur est pris ».
 X suit la loi géométrique de paramètre 0,15.

Conseil

Pour déterminer si une variable aléatoire suit une loi géométrique, on peut commencer par déterminer si elle prend des valeurs entières non nulles.

Calculer $P(X = k)$ ou $P(X \leq k)$ pour X suivant une loi géométrique

Un photographe passionné de nature part réaliser un safari photo autour du monde et s'arrête en Indonésie, dans la région de Sumatra. Il souhaite photographier un orang-outan (*Pongo abelii*) qui vit exclusivement dans cette région. La probabilité qu'il réussisse une photo d'orang-outan est 0,1 chaque jour. Une fois la photo prise, il change de destination.

- 1 Quelle est la probabilité qu'il réussisse une photo le troisième jour ?
- 2 S'il réussit une photo dans les trois premiers jours, il se rendra ensuite à Bornéo pour tenter de photographier d'autres orangs-outans.
Quelle est la probabilité qu'il se rende à Bornéo ?

Solution commentée

- 1 Si on note X la variable aléatoire comptant le nombre de répétitions (donc ici le nombre de jours) nécessaires à la réussite de la photo, X suit une loi géométrique de paramètre $p = 0,1$.
La probabilité de réussir la photo le troisième jour est :
$$P(X = 3) = (1 - p)^{3-1} \times p = 0,081.$$
- 2 La probabilité de réussir la photo dans les trois premiers jours est la somme des probabilités de réussir le premier jour, le deuxième et le troisième. Autrement dit :
$$P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3).$$

Avec $P(X = 2) = (1 - p)^{2-1} \times p = 0,09$ et la question précédente, on trouve que :
$$P(X \leq 3) = 0,271.$$

Pour calculer une probabilité de la forme $P(X \leq k)$, on peut d'abord calculer toutes les probabilités de la forme $P(X = n)$ pour $1 \leq n \leq k$, puis les ajouter.

Le mercredi 6 mai 2020

bcpst Bienaymé-Tchebycheff

Le 6-5-2021

Voir le cours de Johanne Cohen Algorithmes probabilistes

Il y a des remarques vraiment intéressantes « plus fin que Markov »

Theorem (Inégalité de Markov)

Soit X une variable aléatoire à valeurs positive ou nulles.
Alors pour tout $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

- Pour $a > 0$, posons

$$I = \begin{cases} 1 & \text{si } X \geq a, \\ 0 & \text{sinon.} \end{cases}$$

- Puisque $X \geq 0$,

$$I \leq \frac{X}{a}. \quad (3)$$

- Puisque I est une variable à valeur dans $\{0, 1\}$, $E[I] = \Pr(I = 1) = \Pr(X \geq a)$.

- En passant à la moyenne dans (3), on obtient

$$\Pr(X \geq a) = E[I] \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}.$$

Exemple : tirage de pièces

- Considérons une suite de tirage de pièces.

- Posons

$$X_i = \begin{cases} 1 & \text{si la } i\text{ème pièce est pile} \\ 0 & \text{sinon.} \end{cases}$$

- Notons par $X = \sum_{i=1}^n X_i$ le nombre de piles parmi les n tirages.

- On a $E[X] = \sum_{i=1}^n E[X_i] = \frac{n}{2}$.

- L'inégalité de **Markov** donne, pour $\lambda > 0$,

$$\Pr(X \geq \lambda n) \leq \frac{E[X]}{\lambda n} = \frac{n}{2\lambda n} = \frac{1}{2\lambda}.$$

Ou encore

$$\Pr(X \geq \lambda \frac{n}{2}) \leq \frac{1}{\lambda}.$$

Inégalité de Tchebyshev

Lorsque l'on possède une information sur la variance, on peut utiliser l'inégalité suivante, qui est plus fine.

Theorem (Inégalité de Tchebyshev)

Pour tout $a > 0$,

$$\Pr(|X - E[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}.$$

- Observons que l'on a égalité par exemple pour une loi telle que $\Pr(X = a) = 1$.

- Une façon qui peut être plus intuitive de comprendre l'inégalité est d'écrire

$$\Pr(X \geq \mu a) \leq \frac{1}{a},$$

pour tout $a > 0$, où $\mu = E[X]$.

Inégalité de Tchebyshev

Lorsque l'on possède une information sur la variance, on peut utiliser l'inégalité suivante, qui est plus fine.

Theorem (Inégalité de Tchebyshev)

Pour tout $a > 0$,

$$\Pr(|X - E[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}.$$

Remarques

- On a égalité par exemple pour une loi telle que

$$\Pr(X \in \{\mu - a, \mu + a\}) = 1,$$

puisque $E[X] = \mu$, $\Pr(|X - E[X]| \geq a) = 1$, $\text{Var}[X] = 1$.

- Une façon qui peut être plus intuitive de voir l'inégalité est d'écrire

$$\Pr(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2},$$

pour tout $a > 0$, où $\mu = E[X]$, $\sigma^2 = \text{Var}[X]$.

Exemple : tirage de pièces

- Reprenons l'exemple des tirages de pièces.
- Puisque X_i est une variable aléatoire à valeur 0 – 1,

$$E[X_i^2] = \Pr(X_i = 1) = \frac{1}{2}$$

- On a $\text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

- Puisque les X_i sont indépendants,

$$\text{Var}[X] = \sum_{i=1}^n \text{Var}[X_i] = \frac{n}{4}.$$

- L'inégalité de Tchebychev donne pour $\lambda > 0$

$$\begin{aligned} \Pr\left(|X - \frac{n}{2}| \geq \lambda n\right) &\leq \frac{\text{Var}[X]}{\lambda^2 n^2} \\ &\leq \frac{n/4}{\lambda^2 n^2} \\ &\leq \frac{1}{4\lambda^2 n}. \end{aligned}$$

- C'est plus fin que pour Markov.

Inégalité de concentration

On effectue n tirages avec remise d'une carte d'un jeu de 52 cartes.

2. En déduire l'espérance et la variance de la variable aléatoire moyenne $M_n = \frac{1}{n} \sum_{i=1}^n X_i$.

3. Quelle doit-être la valeur minimale de n , pour que la probabilité de s'écarter de l'espérance de plus de 0,1 soit inférieure à 0,05 ?

Dans une classe de 25 élèves, dix sont des garçons.

On effectue n tirages avec remise d'un élève de cette classe pour l'interroger.

Pour le i -ième tirage, on note X_i la variable aléatoire valant 1 si la personne interrogée est un garçon et 0 sinon.

1. Donner l'espérance et la variance de X_i .

2. En déduire l'espérance et la variance de la variable aléatoire moyenne $M_n = \frac{1}{n} \sum_{i=1}^n X_i$.

3. Quelle est la valeur minimale de n , pour laquelle la probabilité de s'écarter de l'espérance d'au moins 0,1 soit inférieure à 0,05 ?

On lance n fois un dé équilibré à six faces et on note le nombre de fois où on obtient la face numéro 4.

Combien de lancers faut-il effectuer au minimum pour que la probabilité de s'écarter de la moyenne d'au moins 0,1 soit inférieure à 0,1 ?

On considère une urne contenant dix boules, dont quatre sont rouges. On tire successivement et avec remise une boule de l'urne et on s'arrête dès que l'on obtient une première boule rouge.

L'objectif est de conjecturer l'espérance de cette loi à l'aide d'un programme informatique.

1. Expliquer l'utilisation d'une boucle while ici.

2. Compléter le programme suivant dont l'objectif est de simuler l'expérience.

from random import randint

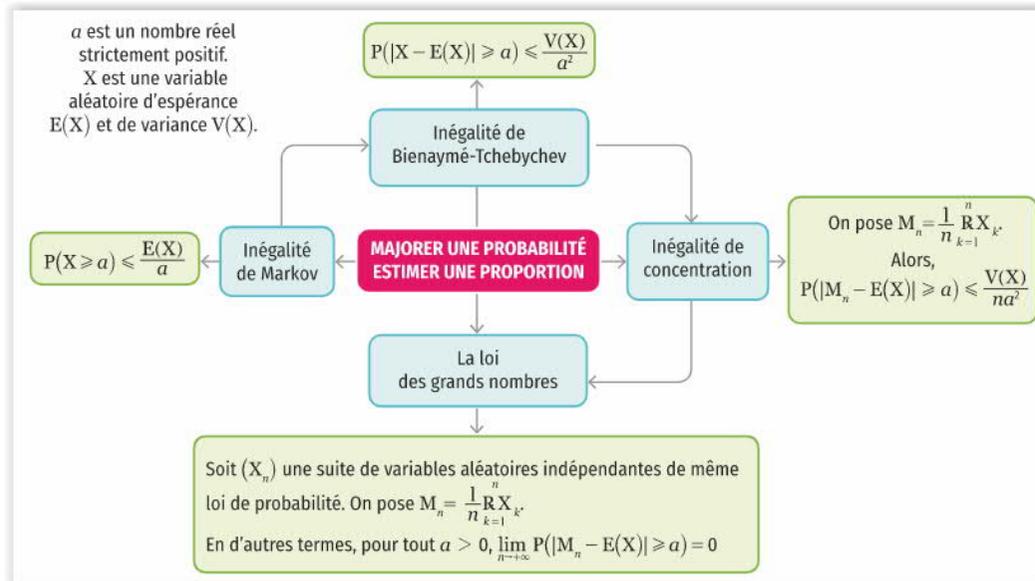
```
def rang(n):
    S = 0
    for i in range(n):
        X = randint(1, 10)
        C = 1
        while X <= ...:
            X = randint(1, 10)
            C = ...
        S = S + C
    return...
```

3°) Quel est le rôle de la variable S ?

Quel est le rôle de la variable S ?

4°) Tester le programme pour différentes grandes valeurs de n. Quelle semble être l'espérance de cette loi ?

Justifier en citant le théorème utilisé.



Histoire des maths :

En probabilités, la loi faible des grands nombres énonce que la moyenne des valeurs d'un échantillon d'une variable aléatoire est proche de l'espérance de cette variable aléatoire lorsque la taille N de l'échantillon est assez grande.

Jacques Bernoulli dans son *Ars Conjectandi* publié en 1713, l'étudie dans le cas de la loi binomiale. Plus tard Siméon Denis Poisson généralise les résultats à d'autres lois de probabilité.

Au milieu du XIXe siècle, Bienaymé (1796-1878), mathématicien français, et Tchebychev (1821-1894) mathématicien russe, énoncent et démontrent l'inégalité qui permettra de prouver de façon rigoureuse cette loi des grands nombres



Le 3 juin 2021

J'avais tapé « Marche aléatoire dans le plan espérance ». J'ai trouvé un extrait de livre aux éditions Dunod.

nombre de lancers nécessaires pour faire apparaître toutes les faces d'un dé cubique

On admet que la valeur exacte de ce nombre moyen est égale à $6 \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right) = 14,7$ [source : .

<https://forums.futura-sciences.com/mathematiques-college-lycee/613379-probabilite-de.html>]. Comparer les résultats obtenus avec cette valeur.

Ce problème s'apparente au « problème du collectionneur ».

Source exercice : Livre « Maths 2de : Exercices résolus – Seconde » de Laurent Darré

24 NOMBRE DE LANCERS | ★★★ | 30 min | P. 35

On souhaite estimer le nombre de lancers nécessaires pour faire apparaître toutes les faces d'un dé à six faces, bien équilibré. Pour cela, on utilise l'algorithme suivant :

```

A, B, C, D, E, F ← 0
N ← 0
Tant que ABCDEF = 0
    Z ← nombre entier aléatoire entre 1 et 6
    Si Z = 1 alors A ← 1
    Si Z = 2 alors B ← 1
    Si Z = 3 alors C ← 1
    Si Z = 4 alors D ← 1
    Si Z = 5 alors E ← 1
    Si Z = 6 alors F ← 1
    N ← N + 1
Fin Tant que
```

a. Expliquer le fonctionnement de cet algorithme

- Expliquer le fonctionnement de cet algorithme
- Modifier cet algorithme pour qu'il simule 1 000 fois l'expérience qui consiste à compter le nombre de lancers nécessaires pour que toutes les faces soient sorties une fois, et qui calcule ensuite le nombre moyen de lancers nécessaires.
- Programmer l'algorithme en langage Python.

Écrire sur la première ligne du programme :

```
from random import randint
```

Cela permet d'utiliser par la suite la fonction `randint(1,6)` qui choisit au hasard un nombre entier entre 1 et 6.

À chaque simulation, les variables A, B, C, D, E et F sont réinitialisées à 0. Le nombre de lancers N est cumulé lors des 1000 simulations. En sortie d'algorithme, une division par 1000 permet de calculer la moyenne.

c. Voici le programme en langage Python :

```
from random import randint
N=0
for i in range(1000):
    A,B,C,D,E,F=0,0,0,0,0,0
    while A*B*C*D*E*F==0:
        Z=randint(1,6)
        if Z==1:
            A=1
        elif Z==2:
            B=1
        elif Z==3:
            C=1
        elif Z==4:
            D=1
        elif Z==5:
            E=1
        else:
            F=1
        N=N+1
print(N/1000)
```

La quatrième ligne est une façon d'attribuer des valeurs à plusieurs variables. L'exécution du programme donne environ 14,7. C'est le nombre moyen de lancers d'un dé pour obtenir toutes les faces au moins une fois.

Le 28-4-2022

<https://forums.futura-sciences.com/mathematiques-college-lycee/613379-probabilite-de.html>

des T_i . Il reste à préciser celles-ci.

Loi de T_i : on a déjà obtenu $i-1$ faces différentes et on lance le dé jusqu'à en obtenir une nouvelle. À chaque lancer, cet événement se réalise avec probabilité $p_i = \frac{\text{nb faces restantes}}{6} = \frac{6-i+1}{6}$, c'est une

expérience de Bernoulli. Dans cette situation, T_i suit la loi géométrique de paramètre p_i . Son espérance est donc $\frac{1}{p_i} = \frac{6}{6-i+1}$.

Finalement, le nombre total moyen de lancers est $\frac{6}{6} + \frac{6}{5} + \dots + \frac{6}{2} + \frac{6}{1} = 14,7$.

C'est le problème du collectionneur.

Le 29-4-2022

collectionneur.py

Créé par [ph-moutou](#)

Créé le 5 juin 2018

Ce programme effectue des tirages aléatoires parmi n objets jusqu'à obtenir la collection complète des objets. L'affichage donne la première et la dernière occurrence ainsi que le nombre d'objets de chaque sorte.

```
from random import randint
def collection(n):
    t=0 #nombre de tirages
    L1,L2,L3=n*[0],n*[0],n*[0]
    while L1.count(0)>0 :
        t+=1
        r=randint(0,n-1)
        if L1[r]==0:L1[r]=t
        L2[r]=t
        L3[r]+=1
    print("{} objets obtenus en {} tirages".format(n,t))
    affichage(L1,L2,L3)
```

```
def affichage(l1,l2,l3):
    print("objet|first|last|nombre")
    for i in range(len(l1)):
        print("{:5}|{:5}|{:4}|{:6}".format(i+1,l1[i],l2[i],l3[i]))
```

```
collection(9)
```

Problème

On effectue une suite de lancers d'une pièce de monnaie. On suppose que les résultats des lancers sont indépendants et que, à chaque lancer, la pièce donne **FACE** avec la probabilité p ($0 < p < 1$) ou **PILE** avec la probabilité $q = 1-p$. L'objet de l'exercice est l'étude du nombre de lancers nécessaires pour obtenir deux FACE de suite, c'est-à-dire lors de deux lancers consécutifs. On suppose donné un espace probabilisé, muni d'une probabilité P , modélisant cette expérience.

• Pour tout entier $n \geq 1$, on note : U_n l'événement : « On obtient deux FACE de suite, pour la première fois aux lancers numéros n et $n+1$ », et l'on pose $u_n = P(U_n)$.

• Pour tout entier $n \geq 2$, on note :

▷ A_n l'événement : « Les n premiers lancers ne donnent pas deux FACE de suite et le $n^{\text{ième}}$ lancer donne FACE. »

▷ B_n l'événement : « Les n premiers lancers ne donnent pas deux FACE de suite et le $n^{\text{ième}}$ lancer donne PILE. »

On pose alors : $x_n = P(A_n)$ et $y_n = P(B_n)$.

5. Informatique.

5.a. Écrire en quelques lignes une fonction `tirage(p)` en langage Python simulant l'expérience et dont l'argument d'entrée est $p \in]0, 1[$ et qui retourne le nombre aléatoire N tel que : « deux FACE de suite ont été obtenus pour la première fois aux rangs N et $N+1$ ».

Pour simuler le lancer de pièce vous utiliserez l'instruction `random()` du module `random` ou l'instruction `rand()` du module `numpy.random`, qui retourne un réel choisi au hasard dans $[0, 1]$.

5.b. Que fait le programme suivant ?

```
def mystere(p, n):
    nbre=10000
    L=[tirage(p) for k in range(nbre)]
    c=0
    for x in L:
        if x==n:
            c+=1
    return c/nbre
```

5.c. Comparer la valeur obtenue en lançant `mystere(2/3, 15)` avec la valeur correspondante obtenue avec votre formule (du DM). Essayez aussi en augmentant la valeur de `nbre`.

Le samedi 12-12-2020

Simulation du lancer d'une pièce telle que la probabilité d'obtenir pile en un lancer soit égale à p

```
from random import random
```

```
def lancer(p):
    r=random()
    if r<=p:
        x=1
    else :
        x=0
    return x
```

DM Sixtine de Leusse avec Mme Lecroël sujet HEC 1998

temps d'attente de deux succès consécutifs pour un schéma de Bernoulli

```
def tirage(p):
    c=0
    u, v=0, 0
    while u*v!=1:
        u, v=lancer(p), u
        c=c+1
    return c
```

On peut remplacer la condition $u*v \neq 1$ par $u+v \neq 2$ ou $(u \neq 1 \text{ or } v \neq 1)$

Le 23-12-2020 en reprenant il m'a semblé qu'il y a une erreur dans la fonction tirage.

Le 23-1-2021

J'ai repris la fonction

```
from random import random
```

```
def lancer(p):
    r=random()
    if r<=p:
        x=1
    else :
        x=0
    return x
```

```
def tps-2succesconsec(p):
    c=1
    u1, u2=lancer(p), lancer(p)
    while u1*u2!=1:
        u1, u2=u2, lancer(p)
        c=c+1
    return c
```

On cherche à simuler la variable X qui est égal au plus petit entier n tel que le n -ième tirage et le $(n+1)$ -ième tirage donnent un succès.

□ `tirage(0.55)`

- 1
- tirage(0.9)
- 2
- tirage(0.3)
- 22
- tirage(0.1)
- 571
-

Le samedi 12-12-2020

J'ai tapé sur Google : BCPST « On n'obtient pas deux FACE de suite ».
 Il s'agit d'un problème cherché avec Sixtine de Leusse.

```
def mystere(p, n):
    nbre=10000
    L=[tirage(p) for k in range(nbre)]
    c=0
    for x in L
        if x==n:
            c=c+1
    return(c/nbre)
```

Une épreuve consiste à faire le produit de 20 nombres aléatoires (loi uniforme sur [0;1]) et à noter le premier chiffre non nul du produit obtenu. Faire des statistiques sur une simulation de 1000 épreuves. Que remarque-t-on ?

-*- coding:Utf-8 -*-

```
from random import*
from math import*

#cette fonction simule une épreuve
def epreuve():
    a=1
    for n in range(20):
        a=a*random()
    a="%e" % (a)      #a devient une chaîne en notation exponentielle
    a=int(a[0])
    return a
```

#####

```
#On crée un dictionnaire
dic={1:0,2:0,3:0,4:0,5:0,6:0,7:0,8:0,9:0}
#On le remplit
for i in range(1000):
    r=epreuve()
    for k in range(1,10):
        if r==k:
            dic[k]+=1
#On lit le dictionnaire
for cle in dic:
    print cle, " ", dic[cle]
```

<https://calque.pagesperso-orange.fr/langages/python/devoirs.html>

<https://calque.pagesperso-orange.fr/math/mantisse.html>

On va dessiner dans ce carré les zones qui correspondent à N=4. On a N=4 si et seulement si le produit xy est dans [4;5[∪ [40;50[ce qui sur le carré correspond à l'ensemble colorié en bleu. Cherchons l'aire de cet ensemble : c'est la somme de quatre intégrales.

$$\int_1^4 \frac{5}{x} - \frac{4}{x} dx + \int_4^5 \frac{5}{x} - 1 dx + \int_4^5 10 - \frac{40}{x} dx + \int_5^{10} \frac{50}{x} - \frac{40}{x} dx$$

On admet que la probabilité de N=4 est égale au quotient de l'aire de l'ensemble colorié par celle du carré ce qui va donner

$$p(N = 4) = \frac{1 + 4 \ln(4) - 5 \ln(5)}{9} + \frac{10 \ln(10)}{81}$$

$$p(N = k) = \frac{1 + k \ln(k) - (k + 1) \ln(k + 1)}{9} + \frac{10 \ln(10)}{81}$$

et plus généralement
 Les valeurs numériques

.k	p(N=k)
1	0,241348
2	0,183209
3	0,145454
4	0,117380
5	0,095007
6	0,076402
7	0,060474
8	0,046549
9	0,034178
	1,0000000

Les valeurs numériques montrent que la loi de N est loin d'être équirépartie.

Produit de trois nombres

Mon excellent camarade Pascal Guelfi a poussé plus loin l'aventure et il a eu l'amabilité de me communiquer la formule suivante qui correspond au produit de trois nombres aléatoires non

$$p(N = k) = \frac{2}{9} - \frac{2k+1}{99} + \frac{1}{9} [(k+1) \ln^2(k+1) - k \ln(k)] - \frac{20 \ln(10)}{81} [(k+1) \ln(k+1) - k \ln(k)] + \frac{110 \ln^2(10)}{729}$$

nuls.

1	0,3105
2	0,1816
3	0,1245
4	0,0937
5	0,0754
6	0,0636
7	0,0556
8	0,0498
9	0,0453
	1,0000

On peut vérifier cette loi en choisissant n=3 sur le formulaire Javascript de la page [vers Benford.](#)

V. Échantillonnage et estimation

1°) J'ai trouvé des documents intéressants le 13 juin 2021 et la veille :

Celui de Kevin Tanguy

Celui de Johan Mathieu

La planche de Galton

Loi binomiale tend vers la loi normale.

Mathématiques, enseignement commun, classe de seconde

Échantillonnage

En liaison avec la partie « Algorithmique et programmation », on définit la notion d'échantillon. L'objectif est de faire percevoir, sous une forme expérimentale, la loi des grands nombres, la fluctuation d'échantillonnage et le principe de l'estimation d'une probabilité par une fréquence observée sur un échantillon.

Connaissances

Échantillon aléatoire de taille n pour une expérience à deux issues.

Version vulgarisée de la loi des grands nombres : « Lorsque n est grand, sauf exception, la fréquence observée est proche de la probabilité. »

Principe de l'estimation d'une probabilité, ou d'une proportion dans une population, par une fréquence observée sur un échantillon.

Expérimentations

Lire et comprendre une fonction Python renvoyant le nombre ou la fréquence de succès dans un échantillon de taille n pour une expérience aléatoire à deux issues.

Observer la loi des grands nombres à l'aide d'une simulation sur Python ou tableur.

Simuler N échantillons de taille n d'une expérience aléatoire à deux issues. Si p est la probabilité d'une issue et f sa fréquence observée dans un échantillon, calculer la proportion des cas où l'écart entre p et f est inférieur ou égal à $1/\sqrt{n}$.

22] Problème HEC 1998 donné à Sixtine de Leusse élève de 2^e année au lycée Hoche en DM pour le 14-2-2021

Le 6-5-2022

Trouvé dans exercices sur variables aléatoires discrètes

Dupuy de Lôme

Exercice 8

On considère une suite $(X_n)_{n \geq 1}$ de variables de Bernoulli indépendantes de même paramètre $p \in]0; 1[$ et l'on étudie la première apparition de deux succès consécutifs dans cette suite.

(a) Montrer qu'il est presque sûr qu'il existe $n \geq 2$ vérifiant $X_n = X_{n-1} = 1$.

(b) On note T la variable aléatoire donnée par $T = \min\{n \geq 2 \mid X_n = X_{n-1} = 1\} \cup \{+\infty\}$.

Calculer $P(T=2)$, $P(T=3)$ et exprimer, pour $n \geq 4$, $P(T=n)$ en fonction de $P(T=n-1)$ et $P(T=n-2)$.

(c) Justifier que T admet une espérance finie et calculer celle-ci.

$$E(T) = (1+p)/p^2$$

On effectue une suite de lancers d'une pièce de monnaie. On suppose que les résultats des lancers sont indépendants et que, à chaque lancer, la pièce donne face avec la probabilité p ($0 < p < 1$) et pile avec la probabilité $q = 1 - p$. L'objet de l'exercice est l'étude du nombre de lancers nécessaires pour obtenir deux faces de suite, c'est à dire lors de deux lancers consécutifs. On suppose donné un espace probabilisé, muni d'une probabilité P , modélisant cette expérience. Pour tout entier $n \geq 1$, on note U_n l'événement : on obtient 2 faces de suite, pour la première fois, aux lancers numéro n et $n+1$, et on pose $u_n = P(U_n)$. Pour tout entier $n \geq 2$, on note A_n l'événement : les n premiers lancers ne donnent pas deux faces de suite et le n -ième lancer B, l'événement : les n premiers lancers ne donnent pas deux faces de suite et le n -ième lancer $y_n = P(B_n)$.

1. a) Déterminer u_1 ; u_2 ; u_3 ; q ; y_3 ; u_3 .

b) Trouver, pour $n \geq 2$, une relation simple entre u_n et u_{n-1} .

c) Pour tout $n \geq 2$ déterminer les probabilités conditionnelles d) En déduire, pour tout $n \geq 2$, les relations de récurrence suivantes : $u_{n+1} = p y_n$ $y_{n+1} = q(y_n + u_n)$.

I. 1°) Peu importe la loi de probabilité de X

Source exercice 21] (distance moyenne de deux points dans un carré) : mathweb.net Stéphane Pasquet

Le 24 mai 2022

distance moyenne de deux points au hasard dans un disque

Source de l'exercice 10] : Stéphane Pasquet <https://www.mathweb.fr/euclide/2021/04/13/une-enigme-pour-moi-help-me-please/> avec Solution de Philippe Rackette

Distance moyenne entre deux points d'un carré

Exercices sur les blocs de 0 et de 1 pris sur la feuille de Albern mathématiques

2°) Exemples

① Une usine produit en moyenne 35 pièces par semaine. On note X la variable aléatoire donnant le nombre de pièces produites par semaine.

Que peut-on dire de la probabilité de que l'usine produise plus de 70 pièces par semaine ?

D'après le contexte, on sait que X est à valeurs positives ou nulles et que $E(X) = 35$ (on assimile l'espérance à la moyenne).

D'après l'inégalité de Markov, $P(X \geq 70) \leq \frac{35}{70}$ soit $P(X \geq 70) \leq \frac{1}{2}$.

$\frac{1}{2}$ est un majorant de $P(X \geq 70)$.

② On lance n fois une pièce non truquée, n étant un entier naturel supérieur ou égal à 1.

On note X la variable aléatoire qui compte le nombre de piles obtenus à l'issue des n lancers.

X suit la loi binomiale de paramètres n (nombre de répétitions) et $\frac{1}{2}$ (probabilité d'un succès).

D'après l'inégalité de Markov, $\forall \lambda \in \mathbb{R}_+^*$ $P(X \geq n\lambda) \leq \frac{E(X)}{n\lambda}$.

Or $E(X) = n \times \frac{1}{2} = \frac{n}{2}$ (formule de l'espérance d'une variable aléatoire qui suit une loi binomiale) donc

$$\frac{E(X)}{n\lambda} = \frac{1}{2\lambda}$$

On en déduit que $\forall \lambda \in \mathbb{R}_+^*$ $P(X \geq n\lambda) \leq \frac{1}{2\lambda}$.

Le lundi 30 mai 2022

<https://fr.quora.com/Quelle-est-la-distance-moyenne-entre-deux-points-choisis-au-hasard-uniformement>

Sébastien Chaumont

Suivre

PhD en Mathématiques appliquées et Probabilité (statistiques), Université Nancy I Henri- Poincaré (France) (Diplôme obtenu en 2002)Vote positif par Sammy Plat

Master 1 Mathématiques, Université de Rennes 1 (2020)[2 ans](#)

Quelle est la distance moyenne entre deux points choisis au hasard (uniformément) sur un cercle ?

On peut trouver la réponse dans ce lien en anglais : mindyourowndecision.com — [What's The Average Distance Of Two Points On A Circle ?](#) donné par [Jeff Peltier dans son commentaire](#) sous une autre réponse. Comme elle est simple et jolie, je la reproduis ici.

$$\int_0^{\pi} \frac{2}{\pi} \sin \frac{x}{2} dx = \frac{4}{\pi}$$

Jacques Dufour

Suivre

Master en Mathématiques et Informatique, Universités (Diplôme obtenu en 1976)[2 ans](#)

[Quelle est la distance moyenne entre deux points choisis au hasard sur un segment ?](#)

Considérons deux variables aléatoires uniformes et indépendantes XX (d'abscisse x) et YY (d'abscisse y) sur un segment $[0;1][0;1]$, leurs abscisses x et y sont également des variables aléatoires indépendantes et uniformément distribuées sur $[0,1].[0,1]$.

La distance $d(X,Y)d(X,Y)$ entre ces deux points est calculée par la formule:

$$d(X,Y)=|x-y|=(x-y)^2 \quad \sqrt{d(X,Y)=|x-y|=(x-y)^2}$$

On obtient la distance moyenne d_{moy} lorsqu'on répète un grand nombre de fois (n) ce calcul d_i (correspondant au choix aléatoire de deux points du segments)

Dans ce cas $d_{moy} = \frac{1}{n} \sum_{i=1}^n d_i$

Si $n \rightarrow \infty$ alors d_{moy} converge vers l'espérance mathématique de cette distance aléatoire, donnée par la formule (espérance d'une variable aléatoire continue)

$$\delta = E[d(X,Y)] = \int_0^1 \int_0^1 |x-y| dx dy = E[d(X,Y)] = \int_0^1 \int_0^1 |x-y| dx dy$$

On utilise un produit de convolution pour réduire le calcul à une intégrale simple en posant $u=x-y.u=x-y$. On obtient alors 2 intégrales de valeurs identiques car on a deux régions selon le signe de u .

Donc on a:

$$\delta = 2 \int_0^1 u^2 du - \int_0^1 (1-u)^2 du = 2 \int_0^1 u^2 (1-u) du$$

Donc

$$\delta = 2 \int_0^1 (u-u^2) du = 2 \left[\frac{u^2}{2} - \frac{u^3}{3} \right]_0^1 = 2 \left(\frac{1}{2} - \frac{1}{3} \right) = \frac{2}{3}$$

Voici la simulation (programme wxMaxima)

```
s:0;for i:1 thru n do
```

```
( ax:random(1.0),bx:random(1.0),
```

```
d:abs(ax-bx),
```

```
s:s+d)$
```

disp(s/n);

0.3333

[Produit de convolution — Wikipédia](#)

Réponse de Sébastien Chaumont:

[Quelle est la distance moyenne entre deux points choisis au hasard \(uniformément\) à l'intérieur d'un carré ?](#)

Jacques Dufour

On répète un grand nombre de fois (n) ce calcul d_i (correspondant au choix aléatoire de deux points du segment).

$$d_{moy} = \frac{1}{n} \sum_{i=1}^n d_i$$

$$d_{moy} \xrightarrow{n \rightarrow +\infty} E(d(X,Y))$$

Simulations :

Génération de nombres au hasard en Python

Pour générer des nombres aléatoires (ou pseudo-aléatoires), on doit importer des fonctions de la bibliothèque random (random signifie hasard en anglais).

Fonctions à connaître

1°) Fonction random()

La fonction `random()` renvoie un réel aléatoire x tel que $0 \leq x < 1$ selon une **loi de distribution uniforme**. Plus précisément, la fonction `random()` génère un nombre à virgule flottante aléatoire (ou plutôt pseudo-aléatoire) de façon uniforme dans l'intervalle $[0;1[$ selon un algorithme qui recrée du hasard artificiellement. Cette fonction est une fonction sans argument.

2°) Fonction randint(a,b)

Cette fonction renvoie un entier relatif x aléatoire tel que $a \leq x \leq b$.

Les arguments a et b doivent être des entiers relatifs tels que $a \leq b$.

3°) Fonction uniform(a,b)

Cette fonction renvoie un réel relatif x aléatoire tel que $a \leq x \leq b$.

La fonction `uniform(a,b)` renvoie un réel aléatoire dans la plage limitée par a et b .

Les arguments a et b doivent être des réels tels que $a \leq b$.

4°) Fonction choice(liste)

La fonction `choice(liste)` renvoie un élément aléatoire d'une liste.

Exemple : La fonction `choice([-1,1])` renvoie -1 et 1 avec la même probabilité.

Principe important

Soit p un réel fixé compris entre 0 et 1.

Si on choisit un réel x au hasard dans l'intervalle $[0; 1[$, la probabilité que $x \leq p$ est égal à p .

3°) Simulation d'une épreuve de Bernoulli telle que la probabilité d'un succès soit égale à p

- Exemple : simulation du lancer d'une pièce telle que la probabilité d'obtenir pile en un lancer soit égale à p

- Programme général :

On code le succès par 1 et l'échec par 0.

On appelle cette fonction `bern(p)` (bern comme « Bernoulli », c'est un choix purement personnel).

Il s'agit d'une fonction prenant pour argument un réel p entre 0 et 1 qui correspond à la probabilité d'un succès.

```
from random import random

def bern(p):
    r=random()
    if r<= p:
        x=1
    else:
        x=0
    return x
```

- Exemple : simulation du lancer d'une pièce truquée telle que la probabilité d'obtenir pile en un lancer soit

égale à $\frac{1}{3}$

On définit une fonction `lancer()` sans argument.

1^{ère} façon : en utilisant la fonction `random`

```
from random import random

def lancer():
    r=random()
    if r<= 1/3:
        x=1
    else:
        x=0
    return x
```

2^e façon : en utilisant la fonction `randint`

```
from random import randint

def lancer():
    r=randint(1, 3)
    if r=1:
        x=1
    else:
        x=0
    return x
```

4°) Simulation d'un schéma de Bernoulli

On considère une épreuve de Bernoulli telle que la probabilité d'un succès soit égale à p .

On répète n fois cette épreuve dans des conditions identiques indépendantes.

On appelle la fonction `si mul (n, p)`. Elle prend pour arguments un entier naturel n supérieur ou égal à 1 et p un réel compris entre 0 et 1.

On utilise la fonction `bern(p)` donc on suppose que celle-ci a été rentrée auparavant.

On utilise une boucle « Pour ».

```
def si mul (n, p):
    X=0
    for i in range(1, n+1):
        x=bern(p)
        X=X+x
    return X
```

On peut remplacer les instructions `x=bern(p)` et `X=X+x` par `X=X+bern(p)`.

5°) Simulation d'une loi géométrique (temps d'attente du premier succès)

On reprend les mêmes notations que précédemment avec une épreuve de Bernoulli que l'on répète dans des conditions identiques indépendantes. On s'intéresse au numéro de l'épreuve qui donne un succès pour la première fois.

On utilise la fonction bern(p) donc on suppose que celle-ci a été rentrée auparavant.
On utilise une boucle « Tantque ».

```
def temps_prem_succes(p):
    T=0
    x=2 # on met n'importe quelle valeur autre que 1 pour que les
    épreuves puissent commencer
    while x !=1:
        x=bern(p)
        T=T+1
    return T
```

T est une variable aléatoire dont la loi sera étudiée plus tard (il s'agit d'une loi géométrique de paramètre p).

VI. Programmes Python de simulations

Corrigé exercice « On dispose d'une pièce non truquée. On étudie deux jeux. »

Le 5 juin 2022

Exercice 18 j'avais noté de changer G en X

Le 23-6 2022

Exercice pour contrôle

Concours blanc ECGB1-Ozenne

8 juin 2022

Concours blanc

4h

Partie C : Application en probabilité

Dans cette partie, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé noté

(Ω, \mathcal{A}, P) .

On considère une urne contenant initialement une boule bleue et une boule rouge. On procède à des tirages successifs d'une boule au hasard selon le protocole suivant :

× si on obtient une boule bleue, on la remet dans l'urne et on ajoute une boule bleue supplémentaire ;

× si on obtient une boule rouge, on la remet dans l'urne et on arrête l'expérience.

On suppose que toutes les boules sont indiscernables au toucher et on admet que l'expérience s'arrête avec une

probabilité égale à 1. On note N la variable aléatoire égale au nombre de boules présentes dans l'urne à la

n

de l'expérience.

11. a) Montrer soigneusement : $\forall n \in \mathbb{N} \setminus \{0, 1\}, P$

$[N = n] =$

$\frac{1}{n(n-1)}$

b) La variable aléatoire N admet-elle une espérance ?

12. Recopier et compléter les lignes incomplètes de la fonction Python suivante de façon à ce qu'elle renvoie une

simulation de la variable aléatoire N.

1 def simuleN():

2 b = 1 // b désigne le nombre de boules bleues dans l'urne

3 while rand() <:

4 b = b+1

5 N =

6 return N

Le 17-7-2022

Exercice pour contrôle

DM2 Frédéric Gaunard 2021-2022 « La première série est de longueur »

Le 26-5-2021

J'avais noté sur une feuille. En bleu, écrit par un élève.

On assimile l'espérance à la moyenne.

Exemple : 35 pièces par semaine donc $E(X)=35$

Sur cette même feuille, j'avais écrit également :

0,9272429954

binomFRép

Je ne sais pas à quoi correspond ce résultat ni à quel calcul.

Le 15-7-2022, je déplace l'original de l'exercice 25 :

25 On fait tourner n fois une roulette à deux secteurs, l'un marqué 1 et l'autre 0. La probabilité d'obtenir le secteur 1 est notée p.

Dans la suite de 1 et de 0 obtenus, on s'intéresse aux « blocs » de 0 et 1. Par exemple, la suite 10010110101 contient 9 blocs : quatre blocs de 0 et cinq blocs de 1.

À chaque suite de n chiffres obtenus, la variable aléatoire S associe le nombre de blocs de la suite.

Par exemple, pour la suite définie plus haut, S prend la valeur $4 + 5 = 9$.

Pour tout entier naturel i tel que $1 \leq i \leq n$, on appelle X_i la variable aléatoire qui, au i-ième tour de roulette, associe le numéro du secteur obtenu.

Pour tout entier naturel i tel que $2 \leq i \leq n$, on appelle Y_i la variable aléatoire telle que $Y_i = 1$ si X_{i-1} est différent de X_i et 0 sinon.

1°) a) Donner les valeurs de $P(X_i = 0)$ et de $P(X_i = 1)$.

b) Justifier que : $P(Y_i = 1) = P(X_{i-1} = 0) \times P(X_i = 1) + P(X_{i-1} = 1) \times P(X_i = 0)$.

c) Déterminer $P(Y_i = 1)$ en fonction de p.

2°) a) Exprimer S en fonction de Y_2, Y_3, \dots, Y_n puis en déduire l'espérance de S en fonction de n et p.

b) Quel nombre moyen de blocs obtient-on pour $n = 1001$ et $p = 0,5$?

3°) On donne le script incomplet d'une fonction Python qui retourne le nombre de blocs dans le cas $n = 1001$ et $p = 0,5$ de la suite définie à la ligne `L=[... for k in range(1001)]`.

```

from random import randint

def blocs01():
    L=[.. for k in range(1001)]
    s=(L[1] !=L[0])
    for i in range(1, 1000) :
        S=S+...
    return (1+s)

```

- a) Recopier ce programme et compléter la ligne $L=[.. \text{ for } k \text{ in } \text{range}(1001)]$.
 b) Quel est le rôle de la ligne $s=(L[1] !=L[0])$
 c) Compléter la ligne $S=S+...$ à l'aide d'un booléen.

Il s'agit du « nombre moyen de séries » dans une suite de pile ou face.

Le lundi 27 juin 2022

Exercices 41 et 42 de la feuille d'exercices Espaces probabilistes sur un univers fini ou infini d'Arnaud Jobin « jouent aux échecs »

Le mercredi 24 mai 2023

Je déplace cette partie du corrigé de l'exercice 5 :

$$P(A) = P(1601 \leq X \leq 1999)$$

Écrit le 26-5-2021 :

On obtient l'affichage 0,92724929954 (commande binomFrép).

Le mercredi 24 mai 2023

5°) Situation d'une suite de pile ou face

On lance n fois une pièce non truquée, n étant un entier naturel supérieur ou égal à 1.

On note X la variable aléatoire qui compte le nombre de piles obtenus à l'issue des n lancers.

X suit la loi binomiale de paramètres n (nombre d'épreuves) et $\frac{1}{2}$ (probabilité d'obtenir pile en un lancer).

On a $E(X) = \frac{n}{2}$ et $V(X) = \frac{n}{4}$.

D'après l'inégalité de Bienaymé-Tchebychev, $\forall \lambda \in \mathbb{R}_+^*$ $P\left(\left|X - \frac{n}{2}\right| \geq \lambda n\right) \leq \frac{1}{4n\lambda^2}$.

2°) Règle du jeu

- Si le numéro obtenu est égal à 1, on gagne 12 €
- Si le numéro obtenu est égal à 2, on perd 6 €
- Si le numéro obtenu est égal à 3, 4, 5 ou 6, on ne gagne ni ne perd rien.

Questions :

- Le jeu est-il favorable ou défavorable au joueur ?
- Quel gain peut-on espérer en moyenne si on joue un très grand nombre de fois ?

Des réponses pourraient être apportées par une simulation. Celle-ci est proposée plus loin dans le cours (VI. 3°).

3°) Notations

On note X le gain algébrique (c'est-à-dire positif ou négatif) du joueur en euros. X est un nombre réel qui peut prendre les 3 valeurs :

$$x_1 = 12$$

$$x_2 = -6$$

$$x_3 = 0$$

Il est légitime de se poser la question de savoir à quoi sert l'introduction de la lettre X .

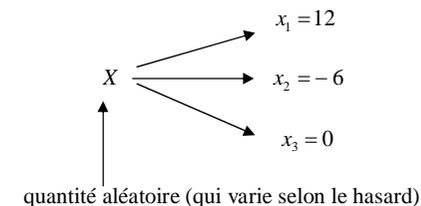
L'introduction de la lettre X se justifie uniquement pour des raisons pratiques qui vont apparaître dans la suite afin de désigner une quantité algébrique aléatoire.

On a tout à fait le droit de décider d'introduire une notation même si on ne voit pas immédiatement à quoi ça sert.

On notera que les différents gains possibles sont notés avec un x minuscule.

On dit que X est une *variable aléatoire réelle* liée à l'expérience aléatoire considérée. L'appellation « variable aléatoire » est très ancienne. Le terme historique est plutôt d'ailleurs « aléa numérique ».

Bilan schématique :



On fera attention aux notations :

X avec une majuscule désigne la variable aléatoire ;

$x_1 = 12$, $x_2 = -6$, $x_3 = 0$ avec des minuscules pour les valeurs que peut prendre la variable aléatoire.

4°) Calculs de probabilités

On munit l'univers des possibles de la loi d'équiprobabilité P .

On va chercher la probabilité de chaque gain possible.

On pose les calculs.

$$\begin{aligned} P(X = x_1) &= P(X = 12) \\ &= P(\text{« obtenir le numéro 1 »}) \\ &= \frac{1}{6} \end{aligned}$$

$$\begin{aligned} P(X = x_2) &= P(X = -6) \\ &= P(\text{« obtenir le numéro 2 »}) \\ &= \frac{1}{6} \end{aligned}$$

$$\begin{aligned} P(X = x_3) &= P(X = 0) \\ &= P(\text{« obtenir le numéro 3, 4, 5 ou 6 »}) \\ &= \frac{4}{6} \\ &= \frac{2}{3} \end{aligned}$$

On a ainsi les probabilités de gain suivantes :

$$\begin{aligned} 12 &\rightarrow \frac{1}{6} \\ -6 &\rightarrow \frac{1}{6} \\ 0 &\rightarrow \frac{2}{3} \end{aligned}$$

On adopte la présentation suivante dans un tableau qui donne la loi de probabilité de X .

x_i	12	-6	0	
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$	Total = 1

Dans notre exemple, la lettre i (comme indice) désigne un entier qui peut prendre 3 valeurs : 1, 2, 3.
 $P(X = x_i)$ désigne la probabilité que X prenne la valeur x_i , c'est-à-dire que le gain algébrique soit égal à x_i .

On écrit $P(X = x_i)$ et non $P(x_i)$ qui ne voudrait rien dire.

X désigne le gain algébrique et non un événement.

X ne représente pas les issues.

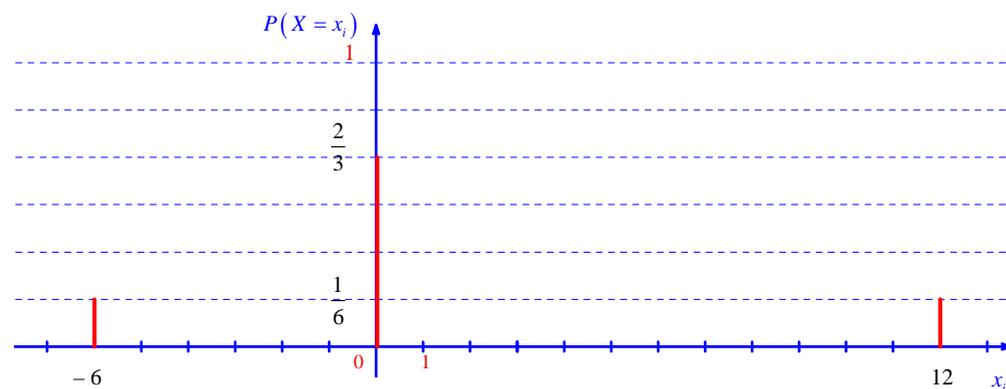
Les issues sont 1, 2, 3, 4, 5, 6.

Les gains possibles sont 12, -6 et 0.

Sans notations particulières, le tableau s'écrirait :

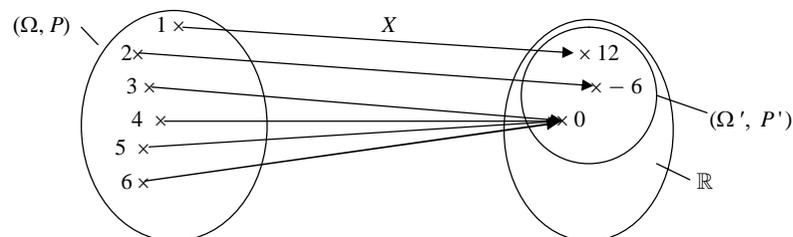
Gain algébrique	12	-6	0	
Probabilité	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$	Total = 1

On peut représenter cette loi de probabilité par un diagramme en bâtons (diagramme en bâtons de la loi de probabilité) comme on le fait en statistiques.



5°) Interprétation

Ce tableau définit une loi de probabilité P' sur l'ensemble $\Omega' = \{12; -6; 0\}$ appelée « loi de probabilité » de X .



Compléments et remarques :

- Il y a 3 valeurs dans Ω' alors qu'il y a 6 valeurs dans Ω .

- X apparaît comme une fonction de Ω dans \mathbb{R} (fonction « entre » Ω et Ω' ou plutôt définie sur Ω à valeurs dans Ω').

- On peut modifier la règle du jeu de manière à avoir 4 valeurs.

Par exemple,

- si le numéro obtenu est égal à 1, on gagne 12 €;
- si le numéro obtenu est égal à 2, on perd 6 €;
- si le numéro obtenu est égal à 3, 4 ou 5, on ne gagne ni ne perd rien ;
- si le numéro obtenu est égal à 6, on gagne 1 €

X peut alors prendre les valeurs $x_1 = 12$, $x_2 = -6$, $x_3 = 0$, $x_4 = 1$.

La loi de probabilité est également changée ; on peut la déterminer très facilement.

II. Définition. Vocabulaire. Conséquence

On va généraliser ce qui a été présenté dans le paragraphe précédent.

1°) Définition

(Ω, P) est un espace probabilisé.

Une **variable aléatoire réelle** est une fonction X définie sur l'univers Ω qui à chaque résultat possible associe un réel.

$$X : \Omega \rightarrow \mathbb{R}$$

Dans la suite, on note x_1, x_2, \dots, x_n les valeurs prises par X (on suppose qu'il y a n valeurs possibles, n étant un entier naturel supérieur ou égal à 1).

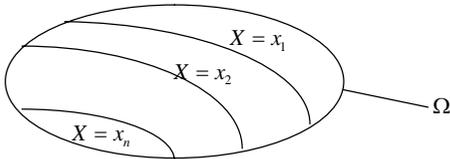
Comme il y a un nombre fini de valeurs, on parle de variable aléatoire discrète (cf. en statistique « caractère discret », « caractère continu »).

Les notations x_i ont déjà été vues en 2° en statistiques avec les effectifs n_i et les fréquences f_i .

2°) Événements élémentaires associés

Pour chaque valeur x_i , l'événement $E_i : \ll X \text{ prend la valeur } x_i \gg$ est noté $(X = x_i)$ (**événement élémentaire associé à x_i**).

3°) Propriété



Les événements $(X = x_1), (X = x_2) \dots$ constituent un « système complet d'événements » de Ω .

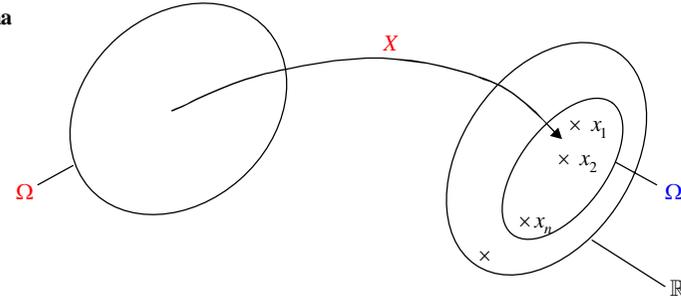
Cela signifie qu'ils sont non vides, deux à deux disjoints et que leur réunion est égale à Ω .

III. Loi de probabilité

1°) Définition

En associant à chaque valeur x_i la probabilité $p_i = P(X = x_i)$, on définit une **loi de probabilité** P' sur l'ensemble $\Omega' = \{x_1, x_2, \dots, x_n\}$ des valeurs prises par X .

Schéma



2°) Présentation

Valeurs possibles de X	x_1	x_2	...	x_n	
Probabilités $p_i = P(X = x_i)$	p_1	p_2	...	p_n	Total = 1

3°) Justification

- $\forall i \in \{1, 2, \dots, n\} \quad 0 \leq p_i \leq 1$
- Les événements $(X = x_1), (X = x_2), \dots, (X = x_n)$ forment un système complet d'événements de Ω donc la somme de leurs probabilités est égale à $P(\Omega) = 1$ soit $p_1 + p_2 + \dots + p_n = 1$. On peut écrire $\sum_{i=1}^{i=n} P(X = x_i) = 1$.

On peut noter que, de manière générale, on a $0 < p_i < 1$ autrement dit que $0 < P(X = x_i) < 1$.
Les inégalités sont strictes.

4°) Notation

Cette loi de probabilité est notée P' ou plutôt P_X (**loi de probabilité de la variable aléatoire X** ou **loi image de P par X**).

5°) Méthode pour déterminer la loi de probabilité d'une variable aléatoire X

- On détermine l'ensemble des valeurs prises par X .

Phrase de rédaction-type :

« X prend les valeurs $x_1 = \dots, x_2 = \dots, \dots$ ».

- On pose les calculs.

$$P(X = x_1) = \dots$$

$$P(X = x_2) = \dots$$

Préciser s'il y a équiprobabilité avant.

- On remplit un tableau (avec les valeurs).

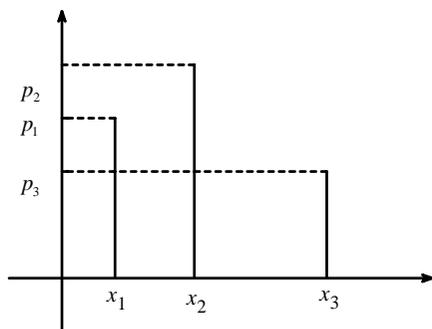
x_i			...		
$P(X = x_i)$...		Total = 1

Dans un tableau de probabilités, les valeurs ne sont pas forcément données dans l'ordre croissant.

6°) Représentation graphique (parfois utilisée)

On représente la loi de probabilité de la variable aléatoire X par un diagramme en bâtons (on suppose que les valeurs sont rangées dans l'ordre croissant).

On peut représenter le diagramme en bâtons (ou plutôt le diagrammes en barres) de la loi de probabilité de la variable aléatoire X .



Ce diagramme permet de visualiser la loi de probabilité.

Utilisation de la calculatrice

On peut visualiser sur l'écran de la calculatrice la loi de probabilité de l'exemple du I (calculatrice TI).

On va utiliser les commandes statistiques.

stats → ÉDIT → 1 : Modifier

Les listes L1, L2 ... sont présentées en colonnes.

Nous n'utiliserons que L1 et L2.

- rentrer les listes en colonnes

Dans la liste L1, on met les valeurs de la variable aléatoire (12 ; -6 ; 0).

On remarquera qu'il n'y a pas besoin de ranger les valeurs dans l'ordre croissant.

Dans la liste L2, on met les probabilités $\left(\frac{1}{6}; \frac{1}{6}; \frac{2}{3}\right)$.

On met les probabilités correspondant aux valeurs ligne à ligne.

On tape bien 1/6, 1/6, 2/3.

L1	L2
12	0.16667
-6	0.16667
0	0,66667

- faire **2nde** **Y=** ou **f(x)** (graph stats)

- choisir 1 : Graph1

- appuyer sur la touche **entrer**

- cliquer sur Aff ou On (calculatrice en anglais ou en français ; Aff signifie « affiché »)

- dans Type, sélectionner le graphique avec rectangles accolés

Liste X : L1

Fréq : L2 (enlever le 1 et mettre L2 grâce à **2nde** **2**)

- appuyer sur la touche **graphe**

Par malchance, si la fenêtre est mal réglée, on obtient un « carré » de « taille » plus ou moins grande (éventuellement un petit « carré »). Il faut adapter la fenêtre.

Par exemple, pour la loi de probabilité de la variable aléatoire considérée dans le I (avec

$P(X = 12) = P(X = -6) = \frac{1}{6}$ et $P(X = 0) = \frac{2}{3}$), on peut choisir la fenêtre suivante :

X min = -12

X max = 14

Xgrad = 1

Y min = -0.1

Y max = 1

Ygrad = 1

On n'obtient pas exactement le diagramme en bâtons de la loi de probabilités mais plutôt le diagramme à barres (diagramme constitué de rectangles non accolés).

IV. Indicateurs d'une variable aléatoire discrète

(Mêmes notations qu'au paragraphe III)

L'ensemble des couples $(x_i, p_i = P(X = x_i))$ peut être vu comme une série statistique dont les valeurs sont les x_i et les fréquences les p_i .

$$\begin{array}{ccc} (x_i, p_i = P(X = x_i)) \\ \uparrow \quad \quad \uparrow \\ \text{valeurs} \quad \quad \text{fréquences} \end{array}$$

On peut donc calculer la moyenne, la variance et l'écart-type (indicateurs de position et de dispersion) comme pour une série statistique.

La moyenne sera appelée « espérance mathématique » conformément au vocabulaire issu des premiers travaux en probabilités qui datent du XVII^e siècle.

Le mot espérance vient du verbe espérer.

Le mot variance vient du verbe varier.

1°) Espérance mathématique

L'espérance mathématique de la variable aléatoire X est définie par la formule :

$$E(X) = \sum_{i=1}^{i=n} x_i \times P(X = x_i) \quad \text{ou} \quad E(X) = \sum_{i=1}^{i=n} (x_i \times P(X = x_i)) \quad (\text{parenthèses pour une meilleure lisibilité})$$

(On peut repasser en rouge les i en indice et dans la somme).

$$E(X) = x_1 \times P(X = x_1) + x_2 \times P(X = x_2) + \dots + x_n \times P(X = x_n)$$

Il s'agit de la moyenne des valeurs x_i pondérée par les probabilités correspondantes (cf. formule de la moyenne d'une série statistique à l'aide des fréquences).

2°) Variance

La notion d'écart moyen (comme en statistiques)

Rappel : L'écart entre deux réels (c'est-à-dire leur distance) est égal à la valeur absolue de leur différence peu importe l'ordre.

Ainsi l'écart entre deux réels a et b est égal à $|a - b|$ ou $|b - a|$.

Comme en statistiques, on pourrait définir l'écart-moyen de X par la formule :

$$|x_1 - E(X)| \times P(X = x_1) + |x_2 - E(X)| \times P(X = x_2) + \dots + |x_n - E(X)| \times P(X = x_n).$$

Quoique cette quantité paraisse naturelle pour mesurer l'écart entre les valeurs et l'espérance, ce n'est pas elle que l'on retient pour mesurer la dispersion autour de l'espérance.

On va utiliser une autre quantité, plus compliquée dans sa définition, mais plus maniable, notamment pour établir des formules.

La variance de la variable aléatoire X est définie par la formule :

$$V(X) = \sum_{i=1}^{i=n} (x_i - E(X))^2 \times P(X = x_i)$$

$$V(X) = (x_1 - E(X))^2 \times P(X = x_1) + (x_2 - E(X))^2 \times P(X = x_2) + \dots + (x_n - E(X))^2 \times P(X = x_n)$$

Il s'agit de la moyenne des carrés des écarts des valeurs x_i par rapport à l'espérance pondérée par les probabilités correspondantes.

De manière immédiate, on a $V(X) \geq 0$.

3°) Écart-type

L'écart-type de la variable aléatoire X est défini par la formule :

$$\sigma(X) = \sqrt{V(X)}$$

L'écart-type est la racine carrée de la variance donc la variance est le carré de l'écart-type : $V(X) = [\sigma(X)]^2$.

4°) Formule de Kœnig-Huygens (démontrée au VIII)

Johann Samuel König (également orthographié Kœnig) est un mathématicien allemand né en 1712 et mort en 1757.

Présenté par Maupertuis, célèbre mathématicien français du XVIII^e siècle, il enseigne les mathématiques à la marquise du Châtelet, et la convertit aux vues de Leibniz, en dépit de Voltaire.

Il devient, vers 1745, professeur de philosophie à Franeker, et en 1749 professeur de philosophie et de droit naturel à La Haye.

Il est correspondant de l'Académie des Sciences de Paris à partir de 1740 et membre de l'Académie de Berlin en 1749.

Christian Huygens (en néerlandais Christiaan Huygens) est un mathématicien, astronome et physicien néerlandais né en 1629 à La Haye (dans les Provinces-Unies) et mort en 1695 dans cette même ville.

Il est proche de Spinoza avec lequel il parle d'optique et d'astronomie, tout en restant éloigné de sa philosophie. Huygens va vivre durant quinze années à Paris. Il est le premier directeur scientifique de l'Académie royale des sciences créée par Colbert en 1666. En 1667, commence la construction de l'Observatoire de Paris (achevée en 1672) par un acte fondateur mémorable : les mathématiciens de l'Académie sous la direction de Huygens tracent le méridien de Paris le jour du solstice d'été, le 21 juin.

À l'automne 1672, il rencontre Gottfried Wilhelm Leibniz dont il devient le professeur de mathématiques et l'ami.

Huygens rencontre Newton le 12 juin 1689 lors de sa troisième visite en Angleterre.

La variance de la variable aléatoire X peut aussi se calculer par la formule :

$$V(X) = \left(\sum_{i=1}^{i=n} (x_i)^2 \times P(X = x_i) \right) - [E(X)]^2$$

5°) Exemple (reprise de l'exemple du I)

Les valeurs prises par X sont $x_1 = 12$, $x_2 = -6$, $x_3 = 0$.

x_i	12	-6	0	
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{4}{6}$	Total = 1

• Espérance mathématique

$$\begin{aligned}
 E(X) &= \sum_{i=1}^{i=3} (x_i \times P(X = x_i)) \\
 &= x_1 \times P(X = x_1) + x_2 \times P(X = x_2) + x_3 \times P(X = x_3) \\
 &= 12 \times \frac{1}{6} + (-6) \times \frac{1}{6} + 0 \times \frac{4}{6} \\
 &= 2 - 1 + 0 \\
 &= 1
 \end{aligned}$$

Dans la somme $\sum_{i=1}^{i=3} \dots$, les nombres 1 et 3 se réfèrent aux indices des valeurs prises par X , que nous avons notées x_1, x_2, x_3 . Le plus petit indice est 1. Le plus grand indice est 3.

Lorsque l'on répète l'expérience aléatoire un très grand nombre de fois, le gain moyen est égal à 1 € (il s'agit d'une limite comme on le verra en Terminale).

Attention à bien interpréter cette valeur de 1 €

1 € n'est pas une valeur possible du gain (les valeurs sont 12, -6 et 0).

Il s'agit d'une valeur moyenne de gains lorsque l'on joue un très grand nombre de fois.

• Variance

Avec la définition

$$\begin{aligned}
 V(X) &= \sum_{i=1}^{i=3} (x_i - E(X))^2 \times P(X = x_i) \\
 &= (12-1)^2 \times \frac{1}{6} + (-6-1)^2 \times \frac{1}{6} + (0-1)^2 \times \frac{4}{6} \\
 &= 11^2 \times \frac{1}{6} + (-7)^2 \times \frac{1}{6} + (-1)^2 \times \frac{4}{6} \\
 &= 121 \times \frac{1}{6} + 49 \times \frac{1}{6} + 1 \times \frac{4}{6} \\
 &= \frac{121}{6} + \frac{49}{6} + \frac{4}{6} \\
 &= \frac{174}{6} \\
 &= 29
 \end{aligned}$$

Avec la formule de Kœnig-Huygens

$$\begin{aligned}
 V(X) &= \left(\sum_{i=1}^{i=3} (x_i)^2 \times P(X = x_i) \right) - [E(X)]^2 \\
 &= 12^2 \times \frac{1}{6} + (-6)^2 \times \frac{1}{6} + 0^2 \times \frac{4}{6} - 1^2 \\
 &= 29
 \end{aligned}$$

Lorsque l'on répète l'expérience aléatoire un très grand nombre de fois, la variance se rapproche de 29.

• Écart-type

$$\begin{aligned}
 \sigma(X) &= \sqrt{V(X)} \\
 &= \sqrt{29}
 \end{aligned}$$

6°) Vocabulaire

On se place dans le cas d'un jeu. On note X le **gain algébrique** du joueur.

L'espérance de X est donc la moyenne des gains pondérés par les probabilités.

- Lorsque $E(X) = 0$, on dit que le jeu est **honnête** ou **équitable**.
- Lorsque $E(X) > 0$, on dit que le jeu est **favorable** au (ou à l'avantage du) joueur.
- Lorsque $E(X) < 0$, on dit que le jeu est **défavorable** au (ou au désavantage du) joueur.

7°) Remarques

- Il est important de mémoriser les gestes associés au tableau donnant la loi de probabilités pour le calcul de l'espérance.

• Pour calculer la variance « à la main », il est préférable d'utiliser la formule de Kœnig-Huygens. Les calculs sont moins fastidieux.

• On démontre aisément que l'espérance d'une variable aléatoire est toujours comprise entre la plus petite et la plus grande valeur prise par la variable aléatoire.

• On démontre aisément que l'écart-type d'une variable aléatoire est toujours inférieur ou égale à la différence entre la plus petite et la plus grande valeur prise par la variable aléatoire.

• On peut utiliser la calculatrice pour calculer l'espérance, la variance et l'écart-type d'une variable aléatoire X .

On procède de la même manière que pour une série statistique « normale » en remplaçant les effectifs par les probabilités. On utilise les listes.

Sur calculatrice TI 83

On appuie sur la touche **stats**.

On utilise EDIT.

1 : Edite

On rentre pour cela les valeurs dans la liste L1 et les probabilités dans la liste L2.

On appuie à nouveau sur la touche **stats**.

On utilise CALC.

1-VarStats L1, L2 ou 1 : Stats 1-Var L1,L2

(la virgule est très importante)

On appuie ensuite sur **entrer**.

L'espérance correspond à la moyenne de la série statistique.

L'écart-type correspond à l'écart-type de la série statistique.

D'autres paramètres que nous n'utiliserons pas cette année apparaissent également

Sur calculatrice TI-83 Plus.fr (modèle noir)

On procède de la même manière sauf pour les calculs.

Après avoir sélectionné CALC, on obtient un écran que l'on complète comme suit :

1-VarStats

List : L1

FreqList : L2

Calculate

On descend sur Calculate puis on appuie alors sur **entrer**.

Sur calculatrice TI-83 Premium-CE

On procède de la même manière sauf pour les calculs.

Après avoir sélectionné CALC, on obtient un écran que l'on complète comme suit :

1-VarStats

XListe : L1

ListFréq : L2

Calculer

On descend sur Calculer puis on appuie alors sur **entrer**.

Attention, on obtient L1 en tapant **2nde** **1** et L2 en tapant **2nde** **2**.

Dans notre exemple introductif,

- on rentre les valeurs possibles de X dans la liste L1 (12, -6, 0) ;

- on rentre les valeurs des probabilités dans la liste L2 ($\frac{1}{6}$, $\frac{1}{6}$, $\frac{2}{3}$).

L1	L2
12	0.16667
-6	0.16667
0	0,66667

On tape bien 1/6, 1/6, 2/3.

On obtient :

$\bar{x} = 1$ (1^{ère} ligne) : il s'agit de l'espérance mathématique (moyenne) ;

$\sigma x = 5,385164807$ (5^e ligne) : il s'agit de l'écart-type.

Sur la 2^e ligne, on lit : $\Sigma x = 1$. Il s'agit de la somme des valeurs. Elle est égale à la moyenne car la somme des coefficients vaut 1.

Sur la 6^e ligne, on lit : $n = 1$ ou $n = 1.00$.

À quoi correspond ce 1 ?

Il correspond au total des probabilités (qui sont dans L2). On sait que la somme des probabilités doit toujours être égale à 1.

En statistiques, lorsque l'on rentre les effectifs dans L2, la valeur de n correspond à l'effectif total de la série (idem pour les fréquences dont la somme vaut 1).

Cela ne nous sert pas. Cela nous permet juste de vérifier que l'on ne s'est pas trompé en rentrant les probabilités.

La calculatrice ne donne pas la variance. Néanmoins, si on en a besoin, on peut élever au carré le résultat de l'écart-type. Attention cependant à bien garder en tête le fait qu'il s'agit de valeurs approchées donc on utilisera davantage la calculatrice pour vérifier.

Avec les listes, il y a un moyen d'obtenir la variance (aller dans listes, puis MATH et sélectionner 8 : variance()).



Pour effacer les listes,

`[stats] > 4 : EffListe > [2nde] [stats] (listes) > L1 ou L2 ...`

8°) Interprétation « concrète » de l'espérance et de la variance d'une variable aléatoire

➤ Le nom « espérance mathématique » est issu du langage des jeux.

L'espérance de gain s'interprète comme la moyenne des gains obtenus en répétant le jeu un grand nombre de fois (voir le paragraphe VII). Le jeu est **favorable** au joueur si son espérance de gain est un nombre strictement positif, et **défavorable** au joueur si c'est un nombre strictement négatif. Lorsque l'espérance est nulle, on dit que le jeu est **équitable**.

L'écart-type du gain mesure la dispersion des gains autour de cette moyenne ; plus il est grand, plus le degré de risque du jeu est grand.

➤ L'espérance d'une variable aléatoire est facile à interpréter.

L'espérance correspond à la moyenne « théorique » si on joue un grand nombre de fois.

Dans l'exemple du I, on a trouvé que l'espérance mathématique de la variable aléatoire X correspondant au gain est de -1 . C'est la moyenne des gains possibles pondérés par les probabilités.

On peut interpréter ce résultat en disant que, si l'on joue un très grand nombre de fois, on va obtenir une moyenne des gains de plus en plus proche de -1 euro. Autrement dit, on perdra 1 euro en moyenne si l'on joue une « infinité » de fois à ce jeu.

➤ La variance ou l'écart-type est plus difficile à interpréter.

La variance ou l'écart-type (l'un est lié à l'autre) mesure la dispersion des valeurs par rapport à l'espérance. L'écart-type fournit une « fourchette » dans laquelle varie la moyenne expérimentale. Ce point, lié à la fluctuation d'une moyenne, ne sera pas développé au niveau du lycée.

Cette année, nous nous contenterons de calculer la variance ou l'écart-type d'une variable aléatoire sans chercher à interpréter le résultat. En revanche, nous nous servirons parfois de l'écart-type pour comparer des jeux ayant la même espérance de gain.

9°) Unités

L'espérance et l'écart-type sont dans la même unité que les valeurs.

La variance est dans l'unité des valeurs au carré.

On n'écrit pas les unités pour des variables aléatoires.

V. Effet d'une transformation affine sur une variable aléatoire

1°) Propriété

X est une variable aléatoire définie sur un univers probabilisé (Ω, P) .

a et b sont deux réels (ce sont des coefficients constants).

On a :

- $E(aX + b) = aE(X) + b$ (linéarité de l'espérance)

- $V(aX + b) = a^2V(X)$

- $\sigma(aX + b) = |a| \sigma(X)$

On notera que les formules de la variance et de l'écart-type ne dépendent pas de b .

Il s'agit de formules analogues à celles sur l'effet d'une fonction affine sur des données (chapitre sur les statistiques).

Utilisation :

On pose $Y = aX + b$. Y est une nouvelle variable aléatoire qui dépend de X .

On suppose que l'on connaît l'espérance, la variance et l'écart-type de X .

Les formules servent trouver l'espérance, la variance et l'écart-type de Y sans refaire tous les calculs, notamment sans refaire le tableau de la loi de probabilité de X .

Des exemples concrets sont donnés dans le paragraphe 5°).

2°) Démonstration

On reprend les notations traditionnelles.

x_i	x_1	x_2	...	x_n	
$P(X = x_i)$	p_1	p_2		p_n	Total = 1

On pose $Y = aX + b$.

Y prend les valeurs $y_1 = ax_1 + b$, $y_2 = ax_2 + b$... $y_n = ax_n + b$ avec les probabilités p_1, p_2, \dots, p_n .
Autrement dit la loi de probabilité de Y est donnée dans le tableau ci-dessous :

y_i	y_1	y_2	...	y_n	
$P(Y = y_i)$	p_1	p_2		p_n	Total = 1

$$\begin{aligned}
 E(Y) &= \sum_{i=1}^{i=n} y_i \times P(Y = y_i) \\
 &= \sum_{i=1}^{i=n} (ax_i + b) \times p_i \\
 &= \sum_{i=1}^{i=n} (ax_i \times p_i + bp_i)
 \end{aligned}$$

$$\begin{aligned}
 &= \left(\sum_{i=1}^{i=n} ax_i \times p_i \right) + \left(\sum_{i=1}^{i=n} b \times p_i \right) \quad (\text{on sépare la somme en deux}) \\
 &= a \underbrace{\left(\sum_{i=1}^{i=n} x_i \times p_i \right)}_{E(X)} + b \underbrace{\left(\sum_{i=1}^{i=n} p_i \right)}_1 \quad (\text{on factorise par } a \text{ et } b \text{ chaque somme})
 \end{aligned}$$

(on « sort » ce qui ne dépend pas de i , c'est-à-dire que l'on utilise la propriété de distributivité pour mettre a et b en facteur dans chaque somme ; si l'on ne voit pas bien, on peut aussi écrire les sommes en extension)

$$= aE(X) + b$$

$$\begin{aligned}
 V(Y) &= \sum_{i=1}^{i=n} (y_i - E(Y))^2 \times P(Y = y_i) \\
 &= \sum_{i=1}^{i=n} (ax_i + b - aE(X) - b)^2 \times p_i \\
 &= \sum_{i=1}^{i=n} (ax_i - aE(X))^2 \times p_i \\
 &= \sum_{i=1}^{i=n} a^2 (x_i - E(X))^2 \times p_i \\
 &= a^2 \sum_{i=1}^{i=n} (x_i - E(X))^2 \times p_i \\
 &= a^2 V(X)
 \end{aligned}$$

$$\begin{aligned}
 \sigma(Y) &= \sqrt{V(Y)} \\
 &= \sqrt{a^2 V(X)} \\
 &= \sqrt{a^2} \times \sqrt{V(X)} \\
 &= |a| \times \sigma(X)
 \end{aligned}$$

Autre démonstration possible pour la variance :

On utilise $V(Y) = E(Y^2) - [E(Y)]^2$ puis on développe.

3°) Exemple

Prenons $a = 3$ et $b = 2$.

Les formules s'écrivent :

$$E(3X + 2) = 3E(X) + 2 \quad (\text{linéarité de l'espérance})$$

$$V(3X + 2) = 9V(X)$$

$$\sigma(3X + 2) = 3\sigma(X)$$

4°) Lien avec les statistiques

- La propriété de linéarité de l'espérance correspond à la propriété de linéarité de la moyenne étudiée en statistiques.
- La propriété sur la variance pour les variables aléatoires correspond à la même propriété pour les séries statistiques.

5°) Mise en application dans des situations concrètes

On applique directement la propriété (on ne refait pas la démonstration).

Situations-types :

On considère une expérience aléatoire.

On note X le gain algébrique en euros pour un jeu associé à cette expérience aléatoire.

On suppose que l'on a déterminé la loi de probabilité de X et que l'on a calculé l'espérance et la variance de X .

On va étudier diverses situations dans lesquelles on modifie les gains.

On note alors Y le nouveau gain (autrement dit, on remplace X par une nouvelle variable aléatoire, donc on change de notation).

On veut obtenir les valeurs de l'espérance et de la variance de Y sans repasser par la loi de probabilité de X .

On ne refait pas le tableau de la loi de probabilité de Y , ce qui présente un gain de temps.

- Situation 1 :

On double tous les gains.

On note Y le nouveau gain en euro.

On a donc $Y = 2X$.

On peut écrire $Y = 2 \times X + 0$.

Pour connaître l'espérance et la variance de Y , on peut appliquer directement la formule avec $a = 2$ et $b = 0$ sans faire de calculs.

On obtient $E(Y) = 2E(X)$ et $V(Y) = 4V(X)$ (le dernier résultat provient du fait que $2^2 = 4$).

- Situation 2 :

On ajoute 3 € à tous les gains.

On note Y le nouveau gain en euro.

On a donc $Y = X + 3$.

On peut écrire $Y = 1 \times X + 3$.

Pour connaître l'espérance et la variance de Y , on peut appliquer directement la formule avec $a = 1$ et $b = 3$ sans faire de calculs.

On obtient $E(Y) = E(X) + 3$ et $V(Y) = V(X)$.

- Situation 3 (situation hybride) :

On double tous les gains et on ajoute 3 € à tous les gains.

On note Y le nouveau gain en euro.

On a donc $Y = 2X + 3$.

Pour connaître l'espérance et la variance de Y , on peut appliquer directement la formule avec $a = 2$ et $b = 3$ sans faire de calculs.

On obtient $E(Y) = 2E(X) + 3$ et $V(Y) = 4V(X)$.

- Situation 4 :

On élève tous les gains au carré.

On note Y le nouveau gain en euro.

On a donc $Y = X^2$.

Cette situation ne rentre pas dans le cadre de la formule.

Il n'y a pas de formule correspondant à cette situation.

Il faut refaire tous les calculs.

VI. Réalité et modèle

1°) Interprétation de l'espérance, variance et écart-type d'une variable aléatoire

On considère une expérience aléatoire \mathcal{E} modélisée par un univers probabilisé (Ω, P) .

On considère une variable aléatoire réelle X définie sur Ω attachée à l'expérience aléatoire.

Pour chaque réalisation de l'expérience, on peut relever la valeur prise par la variable aléatoire X (valeur observée). Chaque valeur dépend du hasard.

Lorsque l'on répète n fois cette expérience aléatoire, on obtient des échantillons de résultats.

On peut calculer la moyenne, la variance et l'écart-type de ces échantillons.

Pour le modèle (Ω, P) , la fréquence de chaque résultat tend vers la probabilité lorsque le nombre n de répétitions augmente.

Donc, lorsque n augmente, la moyenne, la variance et l'écart-type des échantillons se rapprochent de l'espérance mathématique, de la variance et de l'écart-type de X .

Cette remarque permet d'interpréter et de porter une appréciation sur les résultats du calcul des paramètres d'une variable aléatoire.

2°) Fluctuation

La fluctuation des différentes fréquences des résultats entraîne une fluctuation de la moyenne des échantillons autour de l'espérance ainsi que de l'écart-type des échantillons autour de l'écart-type.

Nous admettons que la variance et l'écart-type servent à étudier la fluctuation de la moyenne des échantillons autour de l'espérance (voir chapitre sur les « Moyennes d'échantillons »).

3°) Simulations de variables aléatoires

On peut réaliser des simulations de variables aléatoires à l'aide de tables au hasard ou à l'aide d'outils (calculatrice ou ordinateur).

Nous réaliserons cette année en exercices des simulations de variables aléatoires à l'aide de programmes Python (avec écriture au préalable d'algorithmes de simulation).

Exemple :

On reprend la situation du II.

On lance un dé cubique non truqué dont les faces sont numérotées de 1 à 6.

On note le numéro de la face supérieure.

- Si le numéro obtenu est égal à 1, on gagne 12 €
- Si le numéro obtenu est égal à 2, on perd 6 €
- Si le numéro obtenu est égal à 3, 4, 5 ou 6, on ne gagne ni ne perd rien.

On note X le gain algébrique du joueur en euro.

On définit une fonction gain.

On utilise une structure en « if ...elif ... else ». Rappelons que elif est la contraction de « else et de if ».

```

Fonction gain()
  r prend la valeur d'un entier aléatoire entre 1 et 6
  Si r = 1
    x ← 12
  SinonSi r = 2
    x ← -6
  Sinon
    x ← 0
  Renvoyer x
  
```

On peut également définir une fonction **simulation**(n) qui réalise la simulation de n parties successives indépendantes et qui renvoie le gain moyen (cadre de gauche).
Les n parties permettent d'obtenir un échantillon de taille n (cadre de droite avec utilisation d'une liste).

```

Fonction simulation(n)
  s ← 0
  Pour i allant de 1 à n Faire
    | s ← s + gain( )
  FinPour
  Renvoyer  $\frac{s}{n}$ 
  
```

```

Fonction simulation(n)
  L est une liste vide
  Pour i allant de 1 à n Faire
    | x ← gain()
    | Ajouter x à la liste L
  FinPour
  Renvoyer L
  
```

Voici le programme Python correspondant.

```

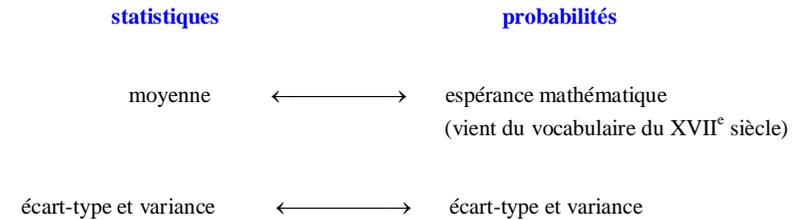
from random import randint

def gain():
    r=randint(1,6)
    if r==1:
        x=12
    elif r==2:
        x=-6
    else:
        x=0
    return x

def simulation(n):
    s=0
    for i in range(1,n+1):
        s=s+gain()
    return s/n
  
```

Le gain moyen est une approximation de l'espérance de X .

4°) Correspondance de termes entre statistiques et probabilités



5°) Interprétation de l'écart-type

On admet le résultat suivant qui sera revu dans le chapitre sur la fluctuation d'échantillonnage d'une moyenne (principe de l'estimation de l'espérance d'une variable aléatoire, ou de la moyenne d'une variable statistique dans une population, par une moyenne observée sur un échantillon).

Inégalité de Bienaymé-Tchebychev (admise sans démonstration cette année, sera démontrée l'année prochaine)

X est une variable aléatoire d'espérance μ et d'écart type σ .

Pour tout réel α strictement positif, on a $P(|X - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$.

Le 3-7-2023

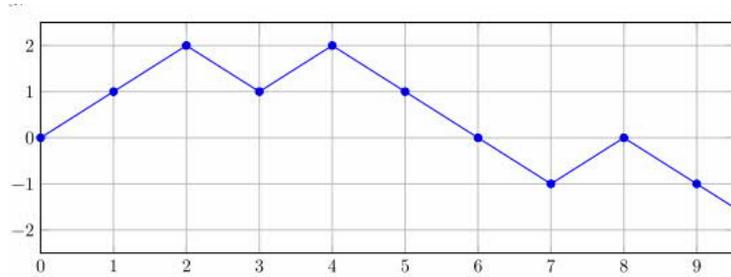
L'espérance d'une variable aléatoire X s'interprète comme la valeur moyenne prise par X lorsqu'on répète un très grand nombre de fois l'expérience.

Par application de cette inégalité pour $\alpha = \sigma$, on obtient $P(|X - \mu| \geq \sigma) \leq 1$, inégalité pas intéressante puisque l'on sait que n'importe quelle probabilité est toujours inférieure ou égale à 1.

Par application de cette inégalité pour $\alpha = 2\sigma$, on obtient $P(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}$.

Par application de cette inégalité pour $\alpha = 3\sigma$,

15



Le jeudi 9 mai 2024

Je déplace le paragraphe suivant :

I. Inégalité de Markov

① Une usine produit

② On lance n fois une pièce non truquée, n étant un entier naturel supérieur ou égal à 1.

On note X la variable aléatoire qui compte le nombre de piles obtenus à l'issue des n lancers.

X suit la loi binomiale de paramètres n (nombre de répétitions) et $\frac{1}{2}$ (probabilité d'un succès).

D'après l'inégalité de Markov, $\forall \lambda \in \mathbb{R}_+^*$ $P(X \geq n\lambda) \leq \frac{E(X)}{n\lambda}$.

Or $E(X) = n \times \frac{1}{2} = \frac{n}{2}$ (formule de l'espérance d'une variable aléatoire qui suit une loi binomiale) donc

$$\frac{E(X)}{n\lambda} = \frac{1}{2\lambda}.$$

On en déduit que $\forall \lambda \in \mathbb{R}_+^*$ $P(X \geq n\lambda) \leq \frac{1}{2\lambda}$.