13

Variables aléatoires, concentration et loi des grands nombres

Les sondages sont omniprésents dans notre actualité quotidienne.
Dans cet extrait, il nous est expliqué que lorsque l'on réalise un sondage, la taille de l'échantillon étudié a de l'influence sur la précision du résultat… et sur le coût du sondage !

Quelle est cette influence et pourquoi ne réalise-t-on pas des sondages sur de plus grands échantillons ?

↪ TP 4 p. 437

VIDÉO
Échantillon
lienmini.fr/maths-s13-01

Pour prendre un bon départ

EXOS
Prérequis
lienmini.fr/maths-s13-02

Les rendez-vous Sésamath

1 Connaître et utiliser la loi de Bernoulli

X suit une loi de Bernoulli de paramètre $p = 0{,}32$.
Donner :
a) $p(X = 1)$. **b)** $p(X = 0)$. **c)** $E(X)$. **d)** $V(X)$.

2 Connaître et utiliser la loi binomiale

On lance 30 fois de suite un dé tétraédrique équilibré.
X est la variable aléatoire donnant le nombre de 4 obtenus sur les 30 lancers.
a) Quelle loi suit X ?
b) Calculer $p(X = 6)$ et $p(X \leqslant 9)$.
c) Calculer la probabilité de faire au moins dix 4 sur les 30 lancers.

3 Calculer l'espérance, la variance et l'écart-type d'une variable aléatoire

X est une variable aléatoire dont la loi de probabilité est donnée dans le tableau suivant.

x_i	-4	7	12	20
$p(X = x_i)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$

Calculer l'espérance $E(X)$, la variance $V(X)$ et l'écart-type $\sigma(X)$ de X.

4 Faire une simulation avec Python

Un jeu consiste à miser 5 euros puis à gagner 50 euros avec une probabilité égale à 0,01 ou gagner 10 euros avec une probabilité égale à 0,04. Dans les autres cas on ne gagne rien.
Recopier et compléter la fonction **Python** suivante pour qu'elle simule le jeu et renvoie le gain algébrique obtenu.

```
import random
def jeu() :
    a=random.random()
    if a<=0.01 :
        k=45
    if a>0.01 and a<=0.05:
        k=...
    if ... :
        k=...
    return k
```

5 Utiliser une valeur absolue

1. x est un nombre réel quelconque. Compléter les phrases suivantes.
a) $|x - 4| \geqslant 3$ traduit le fait que la distance entre x et … est supérieure ou égale à 3.
b) $|x - 5| \leqslant 0{,}5$ traduit le fait que la distance entre x et … est ….
2. Compléter les équivalences suivantes.
a) $2 \leqslant X \leqslant 5 \Leftrightarrow |X - 3{,}5| \leqslant ...$
b) $X < 1\,000$ ou $X > 1\,500 \Leftrightarrow |X - ...| > 250$

Activités

25 min

1 Découvrir la linéarité de l'espérance

On lance un dé équilibré à 6 faces numérotées de 1 à 6.

Un premier jeu permet de gagner 10 euros si le nombre est pair et 5 euros sinon.

Un deuxième jeu permet de gagner 5 euros si le nombre est 1 ou 2 ; de gagner 2 euros si le nombre est 3 et de ne rien gagner sinon.

X est la variable aléatoire donnant le gain au premier jeu et Y est la variable aléatoire donnant le gain au deuxième jeu.

1. Déterminer les lois de probabilité de X et Y.

2. Calculer $E(X)$ et $E(Y)$.

3. On note Z la variable aléatoire donnant la somme des gains des deux jeux : on écrit que $Z = X + Y$.

a) Que gagne-t-on au total si on fait 1 avec le dé ?

b) En considérant les différentes issues du dé, déterminer toutes les valeurs z_i que peut prendre Z et donner la loi de probabilité de Z en recopiant et complétant le tableau ci-dessous.

z_i	5	...	...	...
$p(Z = z_i)$	$\frac{1}{6}$	...	...	...

c) Calculer $E(Z)$ puis comparer $E(Z)$ avec $E(X) + E(Y)$.

4. On double les gains du deuxième jeu (modélisé par la variable aléatoire Y) et on note D la variable aléatoire donnant les gains à ce nouveau jeu.

a) Calculer $E(D)$ après avoir déterminé la loi de probabilité de D.

b) Comparer $E(D)$ et $E(Y)$.

↪ Cours 1 p. 406

Algo 20 min

2 Décomposer une variable aléatoire suivant une loi binomiale

On considère la fonction `binomiale` suivante permettant de simuler une variable aléatoire X en langage **Python**.

```
def binomiale(n,p) :
   s=0
   for i in range(n) :
      s=s+bernoulli(p)
   return s
```

1. Écrire une fonction `bernoulli(p)` permettant de simuler une loi de Bernoulli de paramètre p.

Coup de pouce Ne pas oublier d'importer le module `random`.

2. En la lançant depuis la console, tester cette fonction **Python** pour simuler une variable aléatoire suivant une loi binomiale de paramètres $n = 50$ et $p = 0{,}21$.

3. Expliquer pourquoi la variable aléatoire X suit une loi binomiale de paramètres n et p.

↪ Cours 2 p. 408

25 min

3 Découvrir l'inégalité de Bienaymé-Tchebychev

A ▸ S'éloigner de l'espérance

On considère la variable aléatoire X donnant le gain à un jeu de grattage dont la loi de probabilité est donnée ci-dessous.

x_i	0	2	4	6	10	20	100	1 000	25 000
$p(X = x_i)$	0,682	0,1435	0,103	0,0363	0,0225	0,0126	$9{,}73 \times 10^{-5}$	2×10^{-6}	7×10^{-7}

1. Déterminer $E(X)$.
Pour la suite, on prendra $E(X) = 1{,}423$.

2. Déterminer les probabilités suivantes.
a) $p(|X - E(X)| \geqslant 15)$ **b)** $p(|X - E(X)| \geqslant 100)$ **c)** $p(|X - E(X)| \geqslant 10000)$

3. Que semble-t-on pouvoir dire de la probabilité $p(|X - E(X)| \geqslant \delta)$ quand δ devient grand ?

B ▸ Comparer $p(|X - E(X)| \geqslant \delta)$ et $\dfrac{V(X)}{\delta^2}$

On admet que la variance de X est $V(X) = 450$ (en réalité, 450 est un arrondi à l'entier).

1. Calculer $\dfrac{V(X)}{\delta^2}$ pour $\delta = 15$, pour $\delta = 100$, puis pour $\delta = 10\,000$.

2. Comparer $p(|X - E(X)| \geqslant \delta)$ et $\dfrac{V(X)}{\delta^2}$ dans ces trois cas.

3. Quelle conjecture peut-on faire sur $p(|X - E(X)| \geqslant \delta)$ et sur $\dfrac{V(X)}{\delta^2}$?

↪ Cours 3a p. 410

30 min

4 Découvrir la loi des grands nombres

On considère une pièce de monnaie équilibrée que l'on lance n fois.
Pour tout entier i entre 1 et n, on appelle X_i la variable aléatoire égale à 1 si le résultat du i-ème lancer est PILE et 0 sinon.

1. Déterminer $E(X_i)$ et $V(X_i)$ pour i entier entre 1 et n.

2. Soit $M_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.
Donner une majoration de $p(|M_n - 0{,}5| \geqslant \delta)$ pour δ un réel positif fixé.

3. Soit $\delta = 0{,}1$.
a) Déterminer un entier n_1 à partir duquel on a nécessairement $p(|M_n - 0{,}5| \geqslant \delta) \leqslant 0{,}01$.
b) Déterminer un entier n_2 à partir duquel on a nécessairement $p(|M_n - 0{,}5| \geqslant \delta) \leqslant 0{,}001$.
c) Montrer que $\lim\limits_{n \to +\infty} p(|M_n - 0{,}5| \geqslant \delta = 0)$.
d) Que peut-on en déduire pour la valeur de M_n quand n devient grand ?

4. Reprendre la question **3. c)** avec δ un réel strictement positif quelconque fixé.

5. a) Concrètement, à quoi correspond la variable M_n pour l'échantillon associé aux n lancers de cette pièce de monnaie ?
b) Quelle propriété vue en seconde vient-on d'illustrer dans le cas de cette pièce de monnaie ?

↪ Cours 3c p. 412

Cours

1 Somme de deux variables aléatoires : espérance et variance

Soit X et Y deux variables aléatoires réelles associées à une même expérience aléatoire sur un univers fini Ω.

Définition — Variables aléatoires et opérations

On note :
- **$X + Y$ la variable aléatoire définie sur Ω par $(X + Y)(\omega) = X(\omega) + Y(\omega)$, pour tout $\omega \in \Omega$.**
- **aX la variable aléatoire définie sur Ω par $(aX)(\omega) = a \times X(\omega)$, pour tout $\omega \in \Omega$.**

Exemples

① On tire au sort une adresse dans une ville (l'univers Ω est donc l'ensemble des adresses de cette ville) et on considère les variables aléatoires A et I donnant respectivement le nombre de personnes mAjeures et mIneures habitant à cette adresse.
$A + I$ est donc la variable aléatoire qui, à une adresse, associe la somme des nombres de personnes mAjeures et mIneures y habitant, c'est-à-dire le nombre total de personnes y habitant.

② On lance un dé jaune et un dé vert équilibrés et comportant chacun six faces numérotées de 1 à 6. On note X et Y les variables aléatoires donnant respectivement les résultats affichés par le dé jaune et le dé vert. Très intuitivement, on comprend que $X + Y$ est la variable aléatoire donnant la somme des résultats des deux dés.

Remarque Dans le cadre d'un lancer de trois dés, on peut additionner trois variables aléatoires X, Y et Z, chacune donnant le résultat de chaque dé et obtenir la variable aléatoire $X + Y + Z$.

Propriété — Linéarité de l'espérance

Si X et Y sont deux variables aléatoires sur un même univers fini Ω et a est un nombre réel alors

$$E(X + Y) = E(X) + E(Y) \text{ et } E(aX) = aE(X).$$

Exemple

Dans l'exemple ② précédent, on a $E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = 3,5$ et de même $E(Y) = 3,5$.
Donc $E(X + Y) = E(X) + E(Y) = 3,5 + 3,5 = 7$.
Lorsque l'on lance deux dés, la somme obtenue est en moyenne 7 sur un très grand nombre de lancers.

Remarques
- Si Y est constante égale à b (b un nombre réel), alors $E(X + b) = E(X) + b$.
- Plus généralement on a $E(aX + bY) = aE(X) + bE(Y)$ et en particulier $E(X - Y) = E(X) - E(Y)$.
- Si Z est aussi une variable aléatoire sur Ω, alors $E(X + Y + Z) = E(X) + E(Y) + E(Z)$.

Propriété — Variance

Si X et Y sont deux variables aléatoires sur un même univers fini Ω et a est un nombre réel, alors :
- **$V(X + Y) = V(X) + V(Y)$ si X et Y sont deux variables aléatoires indépendantes (relation d'additivité).**
- **$V(aX) = a^2V(X)$.**
- **$\sigma(aX) = |a|\sigma(X)$.**

Remarque De manière intuitive, deux variables aléatoires sont indépendantes si les résultats de l'une n'ont pas d'influence sur les résultats de l'autre. Pour une définition mathématique, voir l'exercice 149 p. 433.

Exemple

Si X et Y sont deux variables aléatoires sur un même univers fini et indépendantes telles que $V(X) = 3$ et $V(Y) = 7,2$, alors $V(X + Y) = V(X) + V(Y) = 3 + 7,2 = 10,2$ et $\sigma(X + Y) = \sqrt{10,2} \approx 3,19$.
On a également $V(2X) = 4V(X) = 12$.

EXOS Méthodes lienmini.fr/maths-s13-03 Les rendez-vous Sésamath

Exercices résolus

Méthode 1 Calculer une espérance et une variance

Énoncé

X et Y sont deux variables aléatoires indépendantes sur un même univers fini.
La loi de probabilité de X est donnée ci-contre et on a $E(Y) = 2{,}6$ et $V(Y) = 1{,}44$.
Calculer :
a) $E(X + Y)$.
b) $E(2Y)$.
c) $V(X + Y)$.

x_i	0	2
$p(X = x_i)$	0,3	0,7

Solution

a) On a $E(X) = 0 \times 0{,}3 + 2 \times 0{,}7 = 1{,}4$. 1
Donc $E(X + Y) = E(X) + E(Y) = 1{,}4 + 2{,}6 = 4$.

b) $E(2Y) = 2E(Y) = 2 \times 2{,}6 = 5{,}2$.

c) X et Y sont indépendantes donc $V(X + Y) = V(X) + V(Y)$. 2
Or $V(X) = 0{,}3(0 - 1{,}4)^2 + 0{,}7(2 - 1{,}4)^2 = 0{,}84$. 3
Donc $V(X+Y) = 0{,}84 + 1{,}44 = 2{,}28$.

Conseils & Méthodes

1 Si on ne les connaît pas, déterminer d'abord $E(X)$ et $E(Y)$.

2 X et Y doivent être indépendantes pour utiliser la formule $V(X + Y) = V(X) + V(Y)$.

3 Si on ne les connaît pas, déterminer d'abord $V(X)$ et $V(Y)$.

À vous de jouer !

1 X et Y sont deux variables aléatoires telles que $E(X) = 3$, $V(X) = 0{,}5$ et $E(Y) = 5$. Calculer :
a) $E(X + Y)$.
b) $E(2{,}5X)$.
c) $V(2X)$.

2 X et Y sont deux variables aléatoires et indépendantes telles que $E(X) = 15$; $V(X) = 1{,}2$; $E(Y) = 15$ et $V(Y) = 0{,}3$.
Calculer $E(X + Y)$ et $V(X + Y)$.

↪ Exercices 35 à 43 p. 418

Méthode 2 Modéliser avec une somme de variables aléatoires

Énoncé

Une première urne contient trois boules : deux avec le nombre 10 et une avec le nombre – 3.
Une deuxième urne contient sept boules : cinq avec le nombre 3 et deux avec le nombre 0.

On tire une boule dans chaque urne et on regarde la somme des nombres indiqués.

Définir deux variables aléatoires X et Y pour que la somme $X + Y$ modélise la situation.

Solution

On pose X la variable aléatoire qui donne le résultat lors du tirage dans la première urne et Y celle donnant le résultat lors du tirage dans la deuxième urne. 1 Alors le résultat final est donné par $X + Y$.

Conseils & Méthodes

1 Définir des variables aléatoires donnant les résultats à chacune des étapes de l'expérience aléatoire.

À vous de jouer !

3 Au basket, Jean fait dix lancers-francs et Sophie en fait vingt. On s'intéresse au nombre total de lancers-francs réussis par Jean et Sophie. Définir deux variables aléatoires X et Y telles que $X + Y$ modélise la situation.

4 Numa a acheté trois tickets de jeux à gratter différents. On s'intéresse au gain total de Numa.
Proposer une modélisation de cette situation à l'aide d'une somme de plusieurs variables aléatoires.

↪ Exercices 44 à 48 p. 418

Cours

2 Somme de variables aléatoires identiques et indépendantes

Propriété Somme de variables indépendantes suivant une même loi de Bernoulli

Si $X_1, X_2, ..., X_n$ sont des variables aléatoires indépendantes suivant toutes une même loi de Bernoulli de paramètre p, alors $X_1 + X_2 + ... + X_n$ suit une loi binomiale de paramètres n et p.

Démonstration

Pour tout X_i, si l'on considère l'événement « obtenir 1 » de probabilité p comme un succès alors $X_1 + X_2 + ... + X_n$ donne le nombre de succès lorsque l'on réalise n fois de manière indépendante une même expérience de Bernoulli. Donc $X_1 + X_2 + ... + X_n$ suit une loi binomiale de paramètres n et p.

Exemple

Si X_i suit une loi de Bernoulli de paramètre $p = 0{,}17$ pour tout entier i tel que $1 \leqslant i \leqslant 12$ alors $X = X_1 + X_2 + ... + X_{12}$ suit la loi binomiale de paramètres $n = 12$ et $p = 0{,}17$.

Propriété Décomposition d'une variable aléatoire suivant une loi binomiale

Pour X suivant la loi $\mathcal{B}(n\,;p)$, on a $X = X_1 + X_2 + ... + X_n$ où les variables aléatoires $X_1, X_2, ..., X_n$ sont des variables aléatoires indépendantes suivant la loi $\mathcal{B}(p)$.

Exemple

Si X suit la loi binomiale de paramètres $n = 3$ et $p = 0{,}4$ alors $X = X_1 + X_2 + X_3$ où X_1, X_2 et X_3 sont indépendantes et suivent la loi de Bernoulli de paramètre $p = 0{,}4$.

Remarque Cette propriété permet de montrer que si X est une variable aléatoire suivant une loi binomiale de paramètres n et p, alors : $E(X) = np$ et $V(X) = np(1 - p)$.

Définition Échantillon d'une variable aléatoire

Une liste $(X_1\,;X_2\,;...\,;X_n)$ de variables aléatoires indépendantes suivant toutes la même loi est appelée échantillon de taille n associé à cette loi (ou à une variable aléatoire X suivant cette loi).

Exemple

On lance un dé équilibré à six faces : une face porte le nombre 1, deux faces portent le nombre 2 et trois faces portent le nombre 3. Soit X la variable aléatoire donnant le nombre obtenu dont la loi est donnée par le tableau ci-contre.

x_i	1	2	3
$p(X = x_i)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

Un échantillon de taille 4 de la loi suivie par X est la liste $(X_1\,;X_2\,;X_3\,;X_4)$ où chacun des X_i suit la loi de X : cela correspond concrètement à la liste de quatre résultats de quatre lancers du dé.

Propriété Espérance et variance de la somme et de la moyenne d'un échantillon

En considérant un échantillon de taille n $(X_1\,;X_2\,;...\,;X_n)$ d'une variable aléatoire X, et en posant $S_n = X_1 + X_2 + ... + X_n$ et $M_n = \frac{S_n}{n}$, on a :

$$E(S_n) = nE(X) \quad \text{et} \quad V(S_n) = nV(X)$$

$$E(M_n) = E(X) \quad \text{et} \quad V(M_n) = \frac{V(X)}{n}.$$

Remarque M_n (respectivement S_n) est la variable aléatoire correspondant à la moyenne (respectivement la somme) des X_i.

EXOS Méthodes lienmini.fr/maths-s13-03 | Les rendez-vous Sésamath

Exercices résolus

Méthode 3 Sommer des variables aléatoires indépendantes suivant une même loi $\mathscr{B}(p)$

Énoncé

$X_1, X_2, ..., X_{300}$ sont 300 variables aléatoires indépendantes suivant une même loi de Bernoulli de paramètre $p = 0{,}23$.

Quelle loi suit $X = X_1 + X_2 + ... + X_{300}$? En déduire $E(X)$.

Solution

Les X_i étant toutes indépendantes et suivant une loi de Bernoulli de paramètre $p = 0{,}23$, on peut dire que X suit une loi binomiale de paramètres $n = 300$ et $p = 0{,}23$. **1**

Donc $E(X) = n \times p = 300 \times 0{,}23 = 69$. **2**

Conseils & Méthodes

1 S'assurer que toutes les variables sont indépendantes en fonction du contexte.

2 Utiliser la formule du cours directement.

À vous de jouer !

5 On effectue 40 lancers d'une pièce équilibrée. Pour tout entier i entre 1 et 40, X_i est la variable aléatoire donnant 1 si la pièce tombe sur PILE et 0 sinon, au i-ème lancer. Quelle est la loi suivie par $X_1 + X_2 + ... + X_{40}$? À quoi correspond-elle concrètement ?

6 Pour tout entier i tel que $1 \leqslant i \leqslant 8$, X_i suit une loi de Bernoulli de paramètre $p = 0{,}34$. On suppose que toutes les variables aléatoires sont indépendantes. On pose $X = X_1 + X_2 + .. + X_8$. Déterminer $E(X)$.

↪ Exercices 49 à 55 p. 419

Méthode 4 Sommer des variables aléatoires indépendantes suivant une même loi

Énoncé

Une roue de loterie comporte cinq secteurs angulaires égaux. Les deux premiers secteurs valent 300 points, le troisième vaut 100 points et les deux derniers secteurs valent – 400 points. On fait tourner 4 fois de suite cette roue et on gagne la somme de points obtenus lors des 4 lancers de roues.
Z est la variable aléatoire donnant le gain algébrique en points à la fin du jeu.
Décomposer Z en une somme de variables aléatoires identiques et indépendantes puis calculer $E(Z)$.

Solution

1. $Z = X_1 + X_2 + X_3 + X_4$ où X_i est la variable aléatoire donnant le gain algébrique au i-ème lancer de roue pour i entre 1 et 4. **1** Les lancers de roue étant indépendants, les X_i le sont aussi. **2**

X_1, X_2, X_3 et X_4 ont la même loi que la variable aléatoire X donnée ci-contre.

x_i	300	100	– 400
$p(X = x_i)$	0,4	0,2	0,4

2. Tout d'abord $E(X) = 300 \,\square\, 0{,}4 + 100 \,\square\, 0{,}2 + (-400) \,\square\, 0{,}4 = -20$.

On calcule $E(Z) = E(X_1 + X_2 + X_3 + X_4) = 4E(X) = 4 \times (-20) = -80$. **3**

Conseils & Méthodes

1 Repérer la situation qui est répétée.

2 Intuitivement, les actions étant indépendantes, les résultats obtenus le sont aussi.

À vous de jouer !

7 X_1, X_2 et X_3 sont des variables aléatoires indépendantes suivant une même loi binomiale de paramètres $n = 12$ et $p = 0{,}48$. On pose $S = X_1 + X_2 + X_3$.

Calculer $E(S)$.

8 On mise 2 euros puis on lance un dé équilibré à quatre faces. On gagne 3 euros si on fait un 3, et 6 euros si on fait un 4. Sinon on ne gagne rien.
On joue 15 fois à ce jeu. Z est la variable aléatoire donnant le gain algébrique total. Calculer $E(Z)$.

↪ Exercices 56 à 61 p. 419

Cours

3 Concentration et loi des grands nombres

a Inégalité de Bienaymé-Tchebychev

Propriété Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance $E(X) = \mu$ et de variance $V(X) = V$.

Pour tout réel strictement positif δ, on a :

$$p(|X - \mu| \geqslant \delta) \leqslant \frac{V}{\delta^2} \text{ c'est-à-dire } p(X \notin]\mu - \delta\,;\mu + \delta[) \leqslant \frac{V}{\delta^2}.$$

Remarque De manière équivalente, on a $p(|X - \mu| < \delta) = p(X \in]\mu - \delta\,;\mu + \delta[) \geqslant 1 - \frac{V}{\delta^2}$.

Exemple Dans une usine, la variable aléatoire L donnant la largeur en millimètres d'une puce électronique prise au hasard a pour espérance $E(L) = 12$ et pour variance $V(L) = 0{,}01$.

Si la largeur d'une puce n'appartient pas à $]11\,;13[$, c'est-à-dire si $|L - 12| \geqslant 1$, la puce n'est pas commercialisable. La probabilité qu'une puce ne soit pas commercialisable est donc :

$p(|L - 12| \geqslant 1)$ c'est-à-dire $p(|L - \mu| \geqslant \delta)$ avec $\mu = E(L)$ et $\delta = 1$. Comme $V(L) = 0{,}01$, on a $p(|L - 12| \geqslant 1) \leqslant \frac{0{,}01}{1^2}$ d'après l'inégalité de Bienaymé-Tchebychev, or $\frac{0{,}01}{1^2} = 0{,}01$ donc $p(|L - 12| \geqslant 1) \leqslant 0{,}01$.

Propriété Application à $\delta = k\sigma$

Soit X une variable aléatoire d'espérance $E(X) = \mu$, de variance $V(X) = V$ et d'écart-type $\sigma(X) = \sigma$ et $k \in \mathbb{N}^*$.

Les inégalités $p(|X - \mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}$ et $p(|X - \mu| < k\sigma) \geqslant 1 - \frac{1}{k^2}$ sont vérifiées.

Démonstration

On applique l'inégalité de Bienaymé-Tchebychev avec $\delta = k\sigma > 0$ et on obtient :

$$p(|X - \mu| \geqslant k\sigma) \leqslant \frac{V}{(k\sigma)^2} \text{ or } \frac{V}{(k\sigma)^2} = \frac{V}{k^2\sigma^2} = \frac{V}{k^2 V} = \frac{1}{k^2} \text{ donc } p(|X - \mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}.$$

Exemple Dans l'exemple précédent, $\sigma(L) = \sqrt{V} = \sqrt{0{,}01} = 0{,}1$. Ainsi, la probabilité que la largeur de la puce soit éloignée d'au moins $k = 5$ écarts-types, c'est-à-dire $5 \times 0{,}1 = 0{,}5$ de son espérance 12 est inférieure ou égale à $\frac{1}{5^2} = 0{,}04$: il y a au maximum « 4 % de chance » que la largeur d'une puce soit inférieure ou égale à $12 - 0{,}5 = 11{,}5$ mm ou supérieure ou égale à $12 + 0{,}5 = 12{,}5$ mm.

Remarque On mesure donc la dispersion d'une variable aléatoire autour de son espérance en nombre d'écarts-types.

Exemple Pour X qui suit la loi $\mathcal{B}(20\,;0{,}45)$, on a :

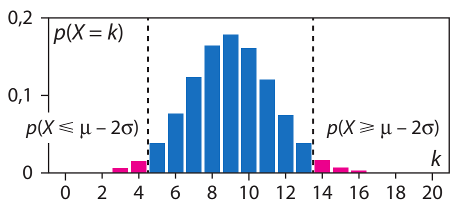
$E(X) = 20 \times 0{,}45 = 9$

et $\sigma(X) = \sqrt{20 \times 0{,}45 \times 0{,}55} \approx 2{,}22$.

Donc, d'après la propriété précédente, on a :

$p(|X - 9| \geqslant 2\sigma(X)) \leqslant \frac{1}{2^2}$ or $\frac{1}{2^2} = 0{,}25$

donc $p(|X - 9| \geqslant 2\sigma(X)) \leqslant 0{,}25$.



D'autre part $p(|X-9| \geqslant 2\sigma(X)) = p(X \leqslant 9 - 2\sigma(X)) + p(X \geqslant 9 + 2\sigma(X)) = p(X \leqslant 4) + p(X \geqslant 14)$ (puisque X ne prend que des valeurs entières) et on observe ci-dessus que $p(X \leqslant 4) + p(X \geqslant 14)$ semble très inférieur à 0,25. Après calculs, on trouve que $p(X \leqslant 4) + p(X \geqslant 14) \approx 0{,}04$: l'inégalité de Bienaymé-Tchebychev donne une **majoration** de $p(|X - \mu| \geqslant \delta)$ par 0,25 qui est **toujours vraie** mais qui est **loin d'être optimale**.

Méthode 5 Utiliser l'inégalité de Bienaymé-Tchebychev

Énoncé

On considère la variable aléatoire D donnant le débit de la Loire en $m^3 \cdot s^{-1}$ à Tours à un instant t.
Une étude statistique permet de considérer que $E(D) = 350$ et $V(D) = 28\,000$.

1. Donner une majoration de $p(|D - 350| \geqslant 200)$ puis interpréter cette majoration dans les termes de l'énoncé.

2. Donner une minoration de la probabilité que le débit de la Loire à Tours à l'instant t soit strictement compris entre 50 et 650 $m^3 \cdot s^{-1}$.

Solution

1. D'après l'inégalité de Bienaymé-Tchebychev, on a :

$p(|D - 350| \geqslant 200) \leqslant \dfrac{28\,000}{200^2} = 0{,}7$. **1**

La probabilité que le débit de la Loire à l'instant t soit inférieur ou égal à $350 - 200 = 150$ $m^3 \cdot s^{-1}$ ou supérieur ou égal à $350 + 200 = 550$ $m^3 \cdot s^{-1}$ est inférieure à 0,7 (ce qui ne donne pas une information très précise). **2**

2. $50 = 350 - 300$ et $650 = 350 + 300$ **3**

On cherche donc :

$p(D \in]50\,;650[) = p(|D - 350| < 300)$

$= p(\overline{|D - 350| \geqslant 300})$

$= 1 - p(|D - 350| \geqslant 300)$. **4**

De plus :

$p(|D - 350| \geqslant 300) \leqslant \dfrac{28\,000}{300^2} = \dfrac{14}{45}$

$\Leftrightarrow -p(|D - 350| \geqslant 300) \geqslant -\dfrac{14}{45}$

$\Leftrightarrow 1 - p(|D - 350| \geqslant 300) \geqslant 1 - \dfrac{14}{45}$

$\Leftrightarrow p(D \in]50\,;650[) \geqslant \dfrac{31}{45}$ avec $\dfrac{31}{45} \approx 0{,}69$.

Remarque On peut aussi retenir directement que $p(|X - \mu| < \delta) \geqslant 1 - \dfrac{V}{\delta^2}$.

Conseils & Méthodes

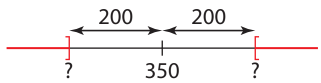
1 Quand on demande une majoration sur une probabilité de la forme $p(|X - \mu| \geqslant \delta)$, on applique l'inégalité de Bienaymé-Tchebychev.

Ici, on a $\mu = 350$ (l'espérance de D), $\delta = 200$ et $V = 28\,000$ (la variance de D).

2 Pour interpréter concrètement, on traduit en langage courant $|D - \mu| \geqslant \delta$.

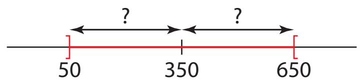
Ici, D est le débit et $|D - 350| \geqslant 200$ veut dire $D \leqslant 350 - 200 = 150$ ou $D \geqslant 350 + 200 = 550$.

On peut représenter la situation sur un axe si nécessaire (ensemble en rouge).



3 Lorsque l'on demande une minoration de la probabilité que la variable aléatoire appartienne à un intervalle contenant μ, on trouve l'écart entre μ et les bornes de cet intervalle et on traduit la probabilité avec la valeur absolue.

On peut représenter la situation sur un axe si nécessaire (ensemble en rouge).



4 On pense ensuite à l'événement contraire.

À vous de jouer !

9 Dans un cabinet médical, le nombre de patient(e)s vu(e)s chaque jour par un médecin est donné par une variable aléatoire P d'espérance $E(P) = 32$ et de variance $V(P) = 9$.

1. a) En appliquant l'inégalité de Bienaymé-Tchebychev, que peut-on dire de $p(|P - 32| < 6)$?
b) Interpréter le résultat dans les termes de l'énoncé.

2. Déterminer une majoration de la probabilité que ce médecin voie soit 21 patient(e)s ou moins soit 41 patient(e)s ou plus en une journée.

10 On considère que le temps passé quotidiennement sur Internet par Luna (en heures) est donné par une variable aléatoire I d'espérance $E(I) = 2$ et de variance $V(I) = 0{,}25$.

Minorer la probabilité que Luna passe entre 1 et 3 heures (exclues) sur Internet aujourd'hui.

↪ Exercices 72 à 78 p. 420

b Inégalité de concentration

Propriété Inégalité de concentration

Soit $(X_1\,;X_2\,;\ldots;X_n)$ un échantillon de variables aléatoires d'espérance μ et de variance V et $M_n=\dfrac{X_1+X_2+\ldots+X_n}{n}$ la variable aléatoire moyenne de cet échantillon.
Pour tout réel strictement positif δ, l'inégalité $p(|M_n-\mu|\geqslant\delta)\leqslant\dfrac{V}{n\delta^2}$ est vérifiée.

Remarque De manière équivalente, on a $p(|M_n-\mu|<\delta)=p(M_n\in\,]\mu-\delta\,;\mu+\delta[)\geqslant 1-\dfrac{V}{n\delta^2}$.

Démonstration

D'après les propriétés sur l'espérance et la variance de la variable aléatoire moyenne d'un échantillon, on a $E(M_n)=\mu$ et $V(M_n)=\dfrac{V}{n}$. En appliquant l'inégalité de Bienaymé-Tchebychev à M_n, on obtient

$p(|M_n-\mu|\geqslant\delta)\leqslant\dfrac{V(M_n)}{\delta^2}$ où $\dfrac{V(M_n)}{\delta^2}=\dfrac{\frac{V}{n}}{\delta^2}=\dfrac{V}{n}\times\dfrac{1}{\delta^2}=\dfrac{V}{n\delta^2}$ c'est-à-dire $p(|M_n-\mu|\geqslant\delta)\leqslant\dfrac{V}{n\delta^2}$.

Exemple On lance n fois un dé équilibré à 8 faces et on nomme X_i la variable aléatoire donnant le résultat du i-ème lancer. On admet que $E(X_i)=4{,}5$ et $V(X_i)=5{,}25$ pour tout entier i entre 1 et n.
Les lancers étant indépendants, $(X_1\,;X_2\,;\ldots;X_n)$ est un échantillon de variables aléatoires d'espérance $\mu=4{,}5$, de variance $V=5{,}25$ et de moyenne $M_n=\dfrac{X_1+X_2+\ldots+X_n}{n}$.
D'après l'inégalité de concentration pour $n=100$ et $\delta=0{,}5$, on a $p(|M_{100}-4{,}5|\geqslant 0{,}5)\leqslant\dfrac{5{,}25}{100\times 0{,}5^2}$

or $\dfrac{5{,}25}{100\times 0{,}5^2}=0{,}21$ donc $p(|M_{100}-4{,}5|\geqslant 0{,}5)$, la probabilité que l'écart entre M_{100} (la moyenne des 100 premiers résultats) et 4,5 soit supérieur ou égal 0,5, est inférieure ou égale à 0,21.

c Loi des grands nombres

Propriété Loi (faible) des grands nombres

Soit $(X_1\,;X_2\,;\ldots;X_n)$ un échantillon de variables aléatoires d'espérance μ et $M_n=\dfrac{X_1+X_2+\ldots+X_n}{n}$ la variable aléatoire moyenne de cet échantillon.
Pour tout réel strictement positif δ fixé, $\lim\limits_{n\to+\infty} p(|M_n-\mu|\geqslant\delta)=0$.

Démonstration

Pour un cas particulier ➥ **Activité 4** p. 405

Exemple On reprend l'exemple précédent et on considère $\delta=0{,}1$.

D'après la loi des grands nombres, $p(|M_n-4{,}5|\geqslant 0{,}1)$, que l'on peut également écrire $p(M_n\notin\,]4{,}4\,;4{,}6[)$, tend vers 0 lorsque la taille de l'échantillon tend vers $+\infty$.

On en déduit que $p(M_n\in\,]4{,}4\,;4{,}6[)$ tend vers 1 lorsque la taille de l'échantillon tend vers $+\infty$. Autrement dit, si l'on fait un nombre suffisamment grand de lancers, on peut rendre l'événement « la moyenne de l'échantillon est dans]4,4 ; 4,6[» aussi probable qu'on le souhaite en prenant n suffisamment grand.

Remarque Dans l'exemple, on aurait pu prendre $\delta=0{,}01$ ou 0,001, etc. : la loi des grands nombres illustre le fait que la moyenne de l'échantillon se rapproche de l'espérance des variables aléatoires quand la taille de l'échantillon « devient grande » comme cela a été vu en Première.

Exercices résolus

Méthode 6 Utiliser l'inégalité de concentration

Énoncé

On considère un échantillon $(X_1 ; X_2 ; \dots ; X_{40})$ de variables aléatoires suivant la loi $\mathcal{B}(10 ; 0{,}9)$ et $M = \dfrac{X_1 + X_2 + \dots + X_{40}}{40}$ la variable aléatoire moyenne de cet échantillon. Donner une majoration de $p(|M - 9| \geqslant 3)$.

Solution

Les $n = 40$ variables aléatoires X_i ont toutes pour espérance $\mu = 10 \times 0{,}9 = 9$ et pour variance $V = 10 \times 0{,}9 \times 0{,}1 = 0{,}9$.

D'après l'inégalité de concentration **1** on a $p(|M - 9| \geqslant 3) \leqslant \dfrac{0{,}9}{40 \times 3^2}$

or $\dfrac{0{,}9}{40 \times 3^2} = \dfrac{0{,}9}{360} = 0{,}002\,5$ donc $p(|M - 9| \geqslant 3) \leqslant 0{,}002\,5$.

Conseils & Méthodes

1 Dans ce contexte d'un échantillon de variables aléatoires, on utilise l'inégalité de concentration en identifiant $\mu = 9$, $\delta = 3$, $n = 40$ et $V = 0{,}9$.

À vous de jouer !

11 100 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.

12 On considère un échantillon $(X_1 ; X_2 ; \dots ; X_{30})$ de 30 variables aléatoires suivant la loi $\mathcal{B}(20 ; 0{,}6)$ et $M = \dfrac{X_1 + X_2 + \dots + X_{30}}{30}$ la variable aléatoire moyenne de l'échantillon.

Justifier que $p(|M - 12| \geqslant 4) \leqslant 0{,}01$.

➥ Exercices 82 à 88 p. 422

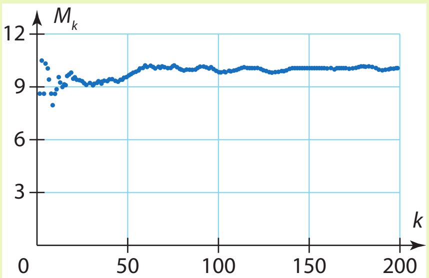
Méthode 7 Utiliser la loi des grands nombres

Énoncé

On considère un échantillon $(X_1 ; X_2 ; \dots ; X_{200})$ de variables aléatoires d'espérance μ et $M_k = \dfrac{X_1 + X_2 + \dots + X_k}{k}$ la variable aléatoire moyenne de l'échantillon $(X_1 ; X_2 ; \dots ; X_k)$ pour k entier entre 1 et 200.

On donne ci-contre le nuage de points $(k ; M_k)$.

Estimer μ.



Solution

On observe que le nuage de points semble « se stabiliser » vers la valeur 10 donc on peut penser que $\mu = 10$.

Conseils & Méthodes

1 D'après la loi des grands nombres, M_k se rapproche de $E(X_i) = \mu$ quand k devient grand.

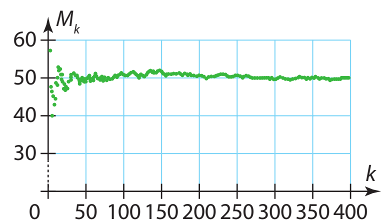
À vous de jouer !

13 On considère un échantillon de variables aléatoires de même loi donnée ci-dessous.

x_i	1	10	30
$p(X = x_i)$	0,4	0,5	0,1

De quelle valeur va se rapprocher la moyenne de cet échantillon quand sa taille augmente ?

14 Reprendre l'énoncé de la Méthode 7 avec le nuage de points suivant (pour un échantillon de taille 400). Estimer μ.



➥ Exercices 89 à 93 p. 423

Exercices résolus

Méthode 8 Utiliser un échantillon dans le cadre d'un prélèvement

↪ Cours 2 p. 408

Énoncé

En France en 2018, selon l'INSEE, 75,4 % des individus de 15 à 29 ans ont réalisé un achat sur Internet au cours des 12 derniers mois.

On interroge 500 personnes de la population française, âgées d'entre 15 et 29 ans, pour savoir si elles ont réalisé un achat sur Internet au cours des 12 derniers mois.

Au vu de la taille de la population en France, on suppose que les tirages au sort successifs ne changent pas les probabilités que la réponse soit positive ou non, et donc que ce prélèvement de 500 personnes par tirage au sort peut être assimilé à un tirage avec remise.

On considère la liste de variables aléatoires $(X_1 ; X_2 ; \dots ; X_{500})$ où X_i vaut 1 si la i-ème personne a réalisé un achat sur Internet au cours de 12 derniers mois et 0 sinon, afin de modéliser l'enquête auprès des 500 personnes.

1. Expliquer pourquoi la liste $(X_1 ; X_2 ; \dots ; X_{500})$ peut être considérée comme un échantillon de variables aléatoires en précisant la loi suivie par les X_i.

2. Soit $S = X_1 + X_2 + \dots + X_{500}$. Calculer $E(S)$ et $V(S)$.

Solution

1. On a supposé que le prélèvement effectué peut être assimilé à un tirage avec remise donc que les X_i sont indépendantes. **1**

D'après l'énoncé, et comme on a précisé que les tirages au sort successifs ne changent pas les probabilités des réponses, les X_i suivent toutes une loi de Bernoulli de même paramètre $p = 0{,}754$.
Ainsi la liste $(X_1 ; X_2 ; \dots ; X_{500})$ peut être considérée un échantillon de variables aléatoires indépendantes et de même loi.

2. Deux méthodes sont possibles.

- $S = X_1 + X_2 + \dots + X_{500}$ donc $E(S) = 500E(X_i)$. Or X_i suit la loi $\mathcal{B}(0{,}754)$ donc $E(S) = 500 \times 0{,}754 = 377$. De plus $V(S) = nV(X_i)$ or X_i suit la loi $\mathcal{B}(p)$ donc $V(S) = 500 \times 0{,}754 \times (1 - 0{,}754) = 92{,}742$. **2**
- Avec la loi binomiale : comme les X_i sont indépendantes et suivent toutes une même loi $\mathcal{B}(0{,}754)$, alors S suit une loi binomiale de paramètres $n = 500$ et $p = 0{,}754$.

$E(S) = np = 500 \times 0{,}754 = 377$ et
$V(S) = np(1 - p) = 500 \times 0{,}754 \times (1 - 0{,}754) = 92{,}742$. **3**

Conseils & Méthodes

1 Comme il a lieu dans une grande population, le tirage, ou le prélèvement, est assimilable à un tirage avec remise donc :
- les variables aléatoires X_i sont indépendantes,
- les variables aléatoires X_i suivent la même loi de Bernoulli.

2 Pour une somme de variables aléatoires X_i identiques et indépendantes, $E(S) = nE(X_i)$ et $V(S) = nV(X_i)$.

3 Comme S suit une loi binomiale de paramètres n et p alors $E(S) = np$ et $V(S) = np(1 - p)$ (voir le chapitre 12).

À vous de jouer !

15 Une urne contient 20 000 billes dont 70 % portent le nombre 1, 25 % le nombre 2 et les autres portent le nombre 9. On effectue un prélèvement de 3 billes. On appelle S la variable aléatoire donnant la somme des trois nombres obtenus.
On suppose que le prélèvement de 3 billes est assimilable à un tirage avec remise.

1. Expliquer pourquoi on peut assimiler ce prélèvement à un tirage avec remise.

2. On décompose S sous forme $S = X_1 + X_2 + X_3$ où X_i donne le nombre porté par la boule lue en i-ème position pour tout entier i entre 1 et 3.
a) Déterminer la loi suivie par X_1 ; X_2 et X_3.
b) Calculer $E(S)$ et $V(S)$.

16 Une entreprise fabrique des vis. On estime que 2 % des vis présente un seul défaut et 1 % présente deux défauts. Les autres ne présentent aucun défaut.
Le service qualité souhaite étudier un échantillon de 200 vis. On pose S la variable aléatoire donnant le nombre de défauts observés sur l'ensemble de l'échantillon.

1. Quel argument peut-on avancer pour justifier que le prélèvement de 200 vis est assimilable à un tirage avec remise dans l'ensemble de la production ?

2. Dans ces conditions, calculer $E(S)$.

↪ Exercices 104 à 106 p. 425

Exercices résolus

Méthode 9 Utiliser l'inégalité de concentration pour trouver la taille d'un échantillon

↪ Cours 3.b p. 412

Énoncé

Une urne contient deux jetons sur lesquels figure le nombre 3, deux jetons sur lesquels figure le nombre 5 et un jeton sur lequel figure le nombre 10.

1. On tire un jeton dans cette urne et on considère la variable aléatoire X donnant le nombre obtenu. Déterminer $E(X)$ et $V(X)$.

2. Combien de tirages avec remise peut-on effectuer dans cette urne pour être sûr au risque de 5 % (ou au seuil de 95 %) que la moyenne des nombres obtenus soit comprise entre 5 et 5,4 exclus ?

Solution

1. $E(X) = \frac{2}{5} \times 3 + \frac{2}{5} \times 5 + \frac{1}{5} \times 10 = 5{,}2$ **1**

$V(X) = \frac{2}{5} \times (3 - 5{,}2)^2 + \frac{2}{5} \times (5 - 5{,}2)^2 + \frac{1}{5} \times (10 - 5{,}2)^2 = 6{,}56$ **2**

2. Les tirages étant avec remise, la liste des variables aléatoires $(X_1 ; \ldots ; X_n)$ donnant les résultats de n tirages successifs est un échantillon de variables aléatoires d'espérance 5,2 et de variance 6,56. Appelons M la variable aléatoire moyenne de cet échantillon. **3**
On cherche à trouver n tel que $p(|M - 5{,}2| \geqslant 0{,}2) \leqslant 0{,}05$. **4**

Comme $p(|M - 5{,}2| \geqslant 0{,}2) \leqslant \frac{6{,}56}{n \times 0{,}2^2}$

d'après l'inégalité de concentration,

c'est à dire que $p(|M - 5{,}2| \geqslant 0{,}2) \leqslant \frac{164}{n}$.

Une condition suffisante pour que $p(|M - 5{,}2| \geqslant 0{,}2) \leqslant 0{,}05$

est $\frac{164}{n} \leqslant 0{,}05$. **5**

On résout $\frac{164}{n} \leqslant 0{,}05 \Leftrightarrow \frac{164}{0{,}05} \leqslant n$ (par positivité de n et 0,05) c'est-à-dire $n \geqslant 3\,280$:

si l'on réalise au moins 3 280 tirages, on est sûr, au risque de 5 %, que la moyenne des résultats obtenus lors de ces tirages sera comprise entre 5 et 5,4.

Remarque Comme $5 = 5{,}2 - 0{,}2$ et $5{,}4 = 5{,}2 + 0{,}2$, on dit que 0,2 est la précision.

Conseils & Méthodes

1 $E(X) = p_1 \times x_1 + \ldots + p_n \times x_n$

2 $V(X) = p_1 \times (x_1 - E(X))^2 + \ldots + p_n \times (x_n - E(X))^2$
On peut aussi utiliser les fonctions statistiques de la calculatrice pour ces deux premiers points.

3 On se place dans les conditions d'application de l'inégalité de concentration en introduisant un échantillon de n variables aléatoires et sa variable aléatoire moyenne.

4 Dire que l'événement $5 < M < 5{,}4$, c'est-à-dire $|M - 5{,}2| < 0{,}2$, est réalisé au risque de 5 % veut dire qu'il y a au plus 5 % « de chance » que $|M - 5{,}2| \geqslant 0{,}2$ autrement dit que $p(|M - 5{,}2| \geqslant 0{,}2) \leqslant 0{,}05$.

5 On travaille par condition suffisante.

À vous de jouer !

17 On lance un dé équilibré à quatre faces numérotées de 1 à 4.
1. On appelle R la variable aléatoire donnant le résultat obtenu. Donner $E(R)$ et $V(R)$.
2. Combien de lancers peut-on effectuer pour être sûr au seuil de 99 % que la moyenne des résultats de ces lancers est comprise entre 2 et 3 exclus ?

18 Le nombre de lignes réalisées à Tetris lors d'une partie par Alexeï est donné par une variable aléatoire d'espérance 125 et de variance 100.
En supposant chacune de ses parties indépendantes, déterminer à partir de combien de parties jouées Alexeï peut être sûr, au risque de 5 %, que sa moyenne de lignes par partie est comprise strictement entre 120 et 130.

↪ Exercices 113 à 117 p. 426

Exercices apprendre à démontrer

VIDÉO
Démonstration
lienmini.fr/maths-s13-04

La propriété à démontrer

Soit n un entier supérieur ou égal à 1 et $p \in]0\,;1[$. Si X est une variable aléatoire suivant une loi binomiale de paramètres n et p, alors $E(X) = np$.

On souhaite démontrer cette propriété en utilisant les prérequis suivants.

① $E(X + Y) = E(X) + E(Y)$ pour toutes variables aléatoires X et Y.

② Si $X_1, X_2, ..., X_n$ sont des variables aléatoires indépendantes suivant toutes une même loi de Bernoulli de paramètre p, alors $X_1 + X_2 + ... + X_n$ suit une loi binomiale de paramètres n et p.

③ Toute variable aléatoire suivant une loi binomiale de paramètres n et p peut être décomposée sous la forme d'une somme de n variables aléatoires indépendantes suivant toutes une même loi $\mathscr{B}(p)$.

Comprendre avant de rédiger

- Des prérequis étant donnés, ils seront certainement utilisés.
- Les prérequis ② et ③ permettent d'utiliser des variables aléatoires suivant des lois de Bernoulli pour construire ou déconstruire une variable aléatoire suivant une loi binomiale : dans la partie hérédité de la démonstration par récurrence, ils permettront de ramener le cas du rang $n + 1$ au cas de rang n et d'utiliser l'hypothèse de récurrence.

Rédiger

La démonstration rédigée

Étape 1 On énonce la proposition à démontrer.

Soit $p \in]0\,;1[$. Pour tout entier $n \geqslant 1$, on considère la propriété $P(n)$: « si X suit $\mathscr{B}(n\,;p)$ alors on a $E(X) = np$ ».

Étape 2 Pour l'initialisation, $n = 1$. Il n'y a qu'une seule répétition.

Initialisation Pour $n = 1$, X suit une loi de Bernoulli de paramètre p. Donc $E(X) = p = 1 \times p$. La propriété est vraie pour $n = 1$.

Étape 3 On rédige l'hypothèse de récurrence.

Hérédité Soit un entier $n \geqslant 1$. Supposons que la propriété $P(n)$ est vraie, c'est-à-dire $E(X) = np$ si X suit $\mathscr{B}(n\,;p)$.
Montrons que P(n + 1) est vraie, c'est-à-dire $E(X) = (n + 1)p$ si X suit $\mathscr{B}(n + 1\,;p)$.
Soit Y une variable aléatoire qui suit $\mathscr{B}(n + 1\,;p)$.

Étape 4
On utilise les prérequis ② et ③ pour écrire Y sous la forme d'une somme d'une variable aléatoire suivant $\mathscr{B}(n\,;p)$ et d'une variable aléatoire suivant $\mathscr{B}(p)$.

Y peut être décomposée en une somme de variables aléatoires indépendantes $Y = X_1 + X_2 + ... + X_n + X_{n+1}$ où les X_i suivent une même loi $\mathscr{B}(p)$.
Alors $X_1 + X_2 + ... + X_n$ suit $\mathscr{B}(n\,;p)$.

Étape 5 On utilise l'hypothèse de récurrence.

Donc $E(X_1 + X_2 + ... + X_n) = np$.
On en déduit que
$$E(Y) = E(X_1 + X_2 + ... + X_n + X_{n+1}) = E(X_1 + X_2 + ... + X_n) + E(X_{n+1})$$
par linéarité de l'espérance.
Donc $E(Y) = np + p$ car X_{n+1} suit une loi de Bernoulli de paramètre p.
Donc $E(Y) = (n + 1)p$.

Étape 5 On écrit une conclusion.

Conclusion On conclut que $P(n)$ est vraie pour tout entier $n \geqslant 1$.
Donc si X suit $\mathscr{B}(n\,;p)$ alors $E(X) = np$ pour tout entier $n \geqslant 1$.

Pour s'entraîner

Soit n un entier naturel supérieur ou égal à 1 et p un réel de $]0\,;1[$. On admettra que si $X_1, ..., X_n, X_{n+1}$ sont indépendantes alors $X_1 + ... + X_n$ et X_{n+1} sont indépendantes.
À l'aide d'un raisonnement par récurrence, démontrer que si X suit une loi binomiale de paramètres n et p alors $V(X) = np(1 - p)$.