

Sommes de variables aléatoires, concentration et loi des grands nombres

COURS • M19 • TERMINALE

Spécialité de terminale, programme officiel. Cours original rédigé avec une IA. Le cours de ton enseignant fait foi.

Sommaire du cours

Ce que tu dois savoir faire	1
Cours	2
Exemples résolus	6
Méthodes et pièges	7
Vers le PASS	7
Exercices	8
Corrigés	9

Ce que tu dois savoir faire

- Définir l'indépendance de deux variables aléatoires et déterminer la loi d'une somme $X + Y$ dans des cas simples.
- Utiliser la linéarité de l'espérance : $E(X + Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$, sans hypothèse d'indépendance.
- Utiliser $V(aX) = a^2V(X)$ et, pour des variables indépendantes, $V(X + Y) = V(X) + V(Y)$.
- Retrouver l'espérance et la variance d'une loi binomiale en l'écrivant comme somme de variables de Bernoulli.
- Définir un échantillon de taille n d'une loi de probabilité et calculer l'espérance, la variance et l'écart type de la somme S_n et de la moyenne M_n .
- Énoncer, démontrer et appliquer l'inégalité de Bienaymé-Tchebychev.
- Appliquer l'inégalité de concentration pour majorer la probabilité d'un écart entre moyenne d'échantillon et espérance.
- Déterminer une taille d'échantillon suffisante pour qu'un écart donné ait une probabilité inférieure à un seuil.
- Énoncer et interpréter la loi faible des grands nombres.

Cours

1. Rappels et notations

Une **variable aléatoire** X sur un univers fini associe un nombre réel à chaque issue. Sa **loi** est la donnée des valeurs $x_1, \dots, x_n$ qu'elle prend et des probabilités $p_i = P(X = x_i)$, de somme 1.

- **Espérance** : $E(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$. C'est la moyenne des valeurs pondérée par leurs probabilités, notée souvent μ .
- **Variance** : $V(X) = E((X - \mu)^2)$ = somme des $p_i (x_i - \mu)^2$. Elle mesure la dispersion de X autour de μ et vérifie la formule de König-Huygens $V(X) = E(X^2) - E(X)^2$.
- **Écart type** : $\sigma(X) = \sqrt{V(X)}$, exprimé dans la même unité que X .

Rappel de Première : pour tous réels a et b , $E(aX + b) = aE(X) + b$ et $V(aX + b) = a^2 V(X)$. Ajouter une constante décale les valeurs sans changer leur dispersion ; multiplier par a multiplie les écarts par $|a|$, donc la variance par a^2 .

2. Somme de deux variables aléatoires et indépendance

Soient X et Y deux variables aléatoires définies sur le même univers. La variable $X + Y$ associe à chaque issue la somme des valeurs prises par X et par Y .

Pour connaître la loi de $X + Y$, il faut en général connaître la **loi du couple** $(X ; Y)$, c'est-à-dire les probabilités $P(X = x \text{ et } Y = y)$. On dit que X et Y sont **indépendantes** si, pour toutes valeurs x et y :

$$P(X = x \text{ et } Y = y) = P(X = x) \times P(Y = y).$$

Intuitivement, connaître la valeur de X ne renseigne pas sur celle de Y . C'est le cas pour les résultats de deux épreuves indépendantes (deux lancers de dé, deux tirages avec remise).

Loi de $X + Y$. Pour une valeur s , $P(X + Y = s)$ est la somme des $P(X = x \text{ et } Y = y)$ pour tous les couples tels que $x + y = s$. Avec deux dés équilibrés X et Y indépendants, chacun des 36 couples a la probabilité $1/36$, et :

s	2	3	4	5	6	7	8	9	10	11	12
$P(X + Y = s)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

La somme n'est pas uniforme : 7 est six fois plus probable que 2, car six couples donnent 7 et un seul donne 2.

3. Linéarité de l'espérance

Théorème. Pour toutes variables aléatoires X et Y sur le même univers et tout réel a :

$$E(X + Y) = E(X) + E(Y) ; E(aX) = aE(X).$$

Ce résultat est vrai **sans hypothèse d'indépendance**.

Justification. On note $p_{xy} = P(X = x \text{ et } Y = y)$. Alors $E(X + Y)$ est la somme des $(x + y)p_{xy}$ sur tous les couples. On sépare en deux sommes. Dans la première, la somme des $x p_{xy}$, on regroupe les termes de même x : la somme des p_{xy} pour y variable vaut $P(X = x)$ (principe des probabilités totales), donc cette première somme vaut $E(X)$. De même, la seconde vaut $E(Y)$.

Par récurrence, pour n variables : $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$.

Exemple : avec deux dés, $E(X) = (1 + 2 + \dots + 6)/6 = 21/6 = 3,5$, donc $E(X + Y) = 7$, sans avoir à utiliser la loi de la somme.

Cette propriété est très puissante : elle permet de calculer l'espérance d'une variable compliquée en la décomposant en une somme de variables simples.

4. Variance d'une somme de variables indépendantes

Théorème. Si X et Y sont indépendantes :

$$V(X + Y) = V(X) + V(Y).$$

Pour n variables $X_1, \dots, X_n$ indépendantes (mutuellement) : $V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n)$.

Justification. Par König-Huygens et linéarité : $V(X + Y) = E((X + Y)^2) - (E(X) + E(Y))^2 = E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2$, soit $V(X + Y) = V(X) + V(Y) + 2[E(XY) - E(X)E(Y)]$. Lorsque X et Y sont indépendantes, $E(XY)$ est la somme des $xy P(X = x)P(Y = y)$, qui se factorise en $E(X) \times E(Y)$. Le crochet est nul.

Conséquences.

- Pour X et Y indépendantes : $V(X - Y) = V(X) + V(-Y) = V(X) + (-1)^2 V(Y) = V(X) + V(Y)$. Les variances s'additionnent même pour une différence : les incertitudes se cumulent.
- Pour X et Y indépendantes et a, b réels : $V(aX + bY) = a^2 V(X) + b^2 V(Y)$.
- Les écarts types ne s'additionnent pas : $\sigma(X + Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}$, qui est inférieur ou égal à $\sigma(X) + \sigma(Y)$.

Attention à l'hypothèse. Sans indépendance, le théorème est faux. Exemple : $V(X + X) = V(2X) = 4V(X)$, et non $2V(X)$. Deux copies indépendantes X_1 et X_2 de même loi donnent $V(X_1 + X_2) = 2V(X)$: additionner des résultats indépendants « disperse moins » que doubler un seul résultat.

5. Application : la loi binomiale

Soit X de loi $B(n, p)$, nombre de succès dans un schéma de Bernoulli. On note X_i la variable valant 1 si la i -ème épreuve est un succès, 0 sinon. Chaque X_i suit la loi de Bernoulli de paramètre p ($E(X_i) = p$, $V(X_i) = p(1 - p)$), les X_i sont indépendantes, et :

$$X = X_1 + X_2 + \dots + X_n.$$

Par linéarité, $E(X) = np$. Par indépendance, $V(X) = np(1 - p)$. On justifie ainsi les formules admises au chapitre sur la loi binomiale, sans aucun calcul de somme de coefficients binomiaux.

Décomposer pour calculer

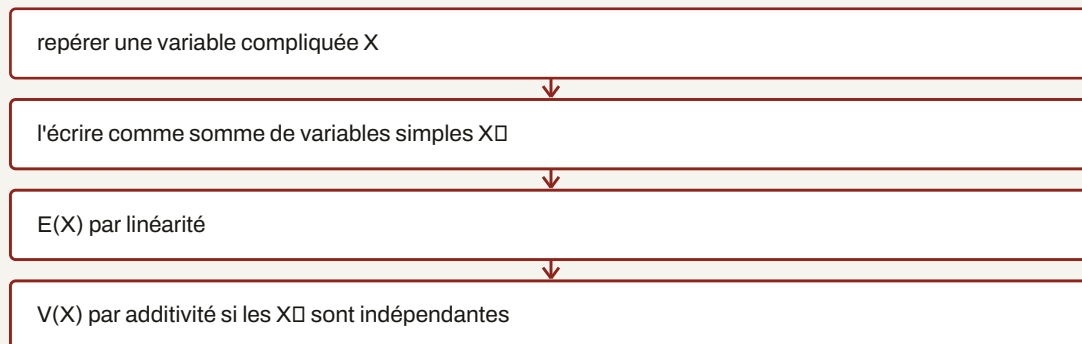


Schéma de principe original • lire les conditions et limites dans le texte

6. Échantillon, somme et moyenne

Définition. Un échantillon de taille n d'une loi de probabilité est une liste $(X_1; X_2; \dots; X_n)$ de n variables aléatoires **indépendantes** suivant toutes **cette même loi**. Il modélise n répétitions indépendantes d'une même expérience (n lancers, n mesures, n individus tirés avec remise).

On note μ l'espérance et V la variance de cette loi ($\sigma = \sqrt{V}$). On définit :

- la **somme** $S_n = X_1 + X_2 + \dots + X_n$;
- la **moyenne** $M_n = S_n/n = (X_1 + \dots + X_n)/n$.

Propriétés.

Variable	Espérance	Variance	Écart type
S_n	$n\mu$	nV	$\sigma\sqrt{n}$
M_n	μ	V/n	$\sigma/\sqrt{n}$

Justification : $E(S_n) = \mu + \dots + \mu = n\mu$ par linéarité ; $V(S_n) = V + \dots + V = nV$ par indépendance. Puis $E(M_n) = E(S_n)/n = \mu$ et $V(M_n) = V(S_n)/n^2 = nV/n^2 = V/n$.

Interprétation. La moyenne d'un échantillon a la même espérance que chaque observation : elle est « centrée » sur μ . Mais son écart type $\sigma/\sqrt{n}$ est **divisé par $\sqrt{n}$** : plus l'échantillon est grand, plus la moyenne se resserre autour de μ . C'est la raison pour laquelle on répète une mesure et on en fait la moyenne. Pour diviser l'écart type par 10, il faut 100 fois plus d'observations.

Cas particulier : si les X_i suivent la loi de Bernoulli de paramètre p , M_n est la **fréquence** de succès, d'espérance p et de variance $p(1 - p)/n$.

7. Inégalité de Bienaymé-Tchebychev

Théorème (inégalité de Bienaymé-Tchebychev). Soit X une variable aléatoire d'espérance μ et de variance V . Pour tout réel δ strictement positif :

$$P(|X - \mu| \geq \delta) \leq V/\delta^2.$$

Démonstration (variable prenant les valeurs $x_1, \dots, x_n$ avec les probabilités $p_1, \dots, p_n$). Par définition, V est la somme des $p_i(x_i - \mu)^2$, qui sont tous positifs ou nuls. En ne gardant que les indices i tels que $|x_i - \mu| \geq \delta$, on obtient une somme plus petite : $V \geq \text{somme, pour } |x_i - \mu| \geq \delta, \text{ des } p_i(x_i - \mu)^2$. Pour ces indices, $(x_i - \mu)^2 \geq \delta^2$, donc : $V \geq \delta^2 \times (\text{somme des } p_i \text{ pour } |x_i - \mu| \geq \delta) = \delta^2 \times P(|X - \mu| \geq \delta)$. On divise par $\delta^2 > 0$ et l'on obtient l'inégalité.

Forme équivalente (par passage au contraire) : $P(|X - \mu| < \delta) \geq 1 - V/\delta^2$.

Avec l'écart type. En prenant $\delta = k\sigma$ ($k > 0$) : $P(|X - \mu| \geq k\sigma) \leq 1/k^2$. Ainsi, pour **toute** variable aléatoire :

- la probabilité de s'écarter de μ d'au moins 2σ est au plus $1/4$;
- celle de s'écarter d'au moins 3σ est au plus $1/9 \approx 0,11$.

Portée et limites. L'inégalité est universelle : elle ne demande que l'espérance et la variance, pas la loi. En contrepartie, elle est souvent **grossière** : la majoration peut être très supérieure à la vraie probabilité (voir l'exemple 2), et elle n'apporte rien si $\delta \leq \sigma$, car alors $V/\delta^2 \geq 1$.

Rappel : $|X - \mu| \geq \delta$ signifie $X \leq \mu - \delta$ ou $X \geq \mu + \delta$, et $|X - \mu| < \delta$ signifie $\mu - \delta < X < \mu + \delta$.

8. Inégalité de concentration

Théorème (inégalité de concentration). Soit M_n la moyenne d'un échantillon de taille n d'une loi d'espérance μ et de variance V . Pour tout réel δ strictement positif :

$$P(|M_n - \mu| \geq \delta) \leq V/(n\delta^2).$$

Démonstration : on applique l'inégalité de Bienaymé-Tchebychev à M_n , d'espérance μ et de variance V/n .

Usage type : taille d'échantillon. Pour que $P(|M_n - \mu| \geq \delta) \leq \alpha$, il suffit que $V/(n\delta^2) \leq \alpha$, c'est-à-dire : $n \geq V/(\alpha\delta^2)$.

Ce n est suffisant, pas forcément nécessaire : comme l'inégalité est grossière, une taille plus petite convient souvent en réalité.

Pour une **fréquence** (loi de Bernoulli de paramètre p inconnu), on utilise la majoration $p(1 - p) \leq 1/4$, valable pour tout p de $[0 ; 1]$: $P(|F_n - p| \geq \delta) \leq 1/(4n\delta^2)$.

9. Loi faible des grands nombres

Théorème (loi faible des grands nombres). Soit $(X_1, \dots, X_n)$ un échantillon d'une loi d'espérance μ , et M_n sa moyenne. Pour tout réel δ strictement positif :

la limite de $P(|M_n - \mu| \geq \delta)$ quand n tend vers $+\infty$ est égale à 0.

Démonstration : par l'inégalité de concentration, $0 \leq P(|M_n - \mu| \geq \delta) \leq V/(n\delta^2)$, et $V/(n\delta^2)$ tend vers 0 quand n tend vers $+\infty$. Le théorème des gendarmes conclut.

Interprétation. Quelle que soit la précision δ choisie, la probabilité que la moyenne de l'échantillon s'écarte de μ d'au moins δ devient aussi petite que l'on veut quand la taille de l'échantillon augmente. En particulier, la **fréquence** d'un événement dans n répétitions indépendantes se rapproche de sa **probabilité** : c'est ce qui justifie l'estimation d'une probabilité par une fréquence observée, et la simulation (méthode de Monte-Carlo).



Ce que dit et ne dit pas la loi des grands nombres

elle concerne la moyenne de n observations indépendantes

l'écart à μ devient improbable quand n augmente

elle ne prédit aucun résultat individuel

les tirages passés ne compensent pas les suivants

Repères complémentaires, sans séquence causale

Contresens à éviter. La loi des grands nombres ne dit pas que le hasard « compense » : après 10 piles consécutifs avec une pièce équilibrée, le lancer suivant donne toujours face avec probabilité $1/2$. La fréquence se rapproche de $1/2$ parce que les 10 piles initiaux pèsent de moins en moins dans une moyenne sur un nombre croissant de lancers, non parce que des faces viendraient « rattraper ».

Exemples résolus

Exemple 1 : somme et différence de deux dés

On lance deux dés équilibrés, de résultats X et Y (indépendants). a) Calculer $E(X)$ et $V(X)$. b) En déduire l'espérance et la variance de $S = X + Y$ et de $D = X - Y$. c) Vérifier $E(S)$ à l'aide de la loi de S .

Résolution. a) $E(X) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 21/6 = 3,5$. $E(X^2) = (1 + 4 + 9 + 16 + 25 + 36)/6 = 91/6$. Par König-Huygens : $V(X) = 91/6 - 3,5^2 = 91/6 - 49/4 = 182/12 - 147/12 = 35/12 \approx 2,92$. b) Par linéarité : $E(S) = 3,5 + 3,5 = 7$ et $E(D) = 3,5 - 3,5 = 0$. X et Y sont indépendantes, donc $V(S) = 35/12 + 35/12 = 35/6 \approx 5,83$ et $V(D) = V(X) + V(Y) = 35/6$ également (les variances s'additionnent aussi pour une différence). c) Avec le tableau du cours : $E(S) = (2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6 + 8 \times 5 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1)/36 = (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12)/36 = 252/36 = 7$. Le résultat est le même, mais la linéarité l'obtient sans calcul de loi.

Exemple 2 : Bienaymé-Tchebychev face au calcul exact

X suit la loi $B(100 ; 0,5)$, nombre de piles en 100 lancers d'une pièce équilibrée. a) Calculer $E(X)$ et $V(X)$. b) À l'aide de l'inégalité de Bienaymé-Tchebychev, majorer $P(|X - 50| \geq 10)$. c) La calculatrice donne $P(X \leq 40) + P(X \geq 60) \approx 0,0569$. Commenter. d) En déduire un minorant de $P(40 < X < 60)$.

Résolution. a) $E(X) = 100 \times 0,5 = 50$; $V(X) = 100 \times 0,5 \times 0,5 = 25$ ($\sigma = 5$). b) Avec $\delta = 10$: $P(|X - 50| \geq 10) \leq 25/10^2 = 0,25$. c) L'événement $|X - 50| \geq 10$ est $X \leq 40$ ou $X \geq 60$, de probabilité exacte $\approx 0,0569$. La majoration 0,25 est **vraie mais environ quatre fois trop large** : l'inégalité, qui n'utilise que la variance, est grossière. d) Par passage au contraire : $P(|X - 50| < 10) \geq 1 - 0,25 = 0,75$, c'est-à-dire $P(40 < X < 60) \geq 0,75$ (la vraie valeur est environ 0,943).

Exemple 3 : taille d'un sondage

On veut estimer la proportion p inconnue d'électeurs favorables à un projet en interrogeant n personnes au hasard (tirages assimilés à des tirages avec remise). On note F_n la fréquence de réponses favorables. a) Donner $E(F_n)$ et majorer $V(F_n)$ indépendamment de p . b) Déterminer une taille n suffisante pour que $P(|F_n - p| \geq 0,05) \leq 0,05$.

Résolution. a) F_n est la moyenne d'un échantillon de taille n de la loi de Bernoulli de paramètre p : $E(F_n) = p$ et $V(F_n) = p(1 - p)/n$. Comme $p(1 - p) = 1/4 - (p - 1/2)^2 \leq 1/4$, on a $V(F_n) \leq 1/(4n)$. b) Par l'inégalité de concentration : $P(|F_n - p| \geq 0,05) \leq V(F_n)/0,05^2 \leq 1/(4n \times 0,0025) = 1/(0,01n) = 100/n$. Il suffit que $100/n \leq 0,05$, soit $n \geq 100/0,05 = 2\,000$ personnes. C'est une taille suffisante ; une méthode plus fine (approximation normale) montre qu'environ 400 personnes suffisent déjà, ce qui illustre à nouveau le caractère grossier de l'inégalité.

Méthodes et pièges

- Pour une espérance, décompose la variable en somme de variables simples et applique la linéarité : aucune indépendance n'est nécessaire.
- Pour une variance de somme, vérifie et écris l'indépendance avant d'additionner les variances.
- $V(X - Y) = V(X) + V(Y)$ pour des variables indépendantes : jamais de soustraction de variances.
- Ne confonds pas $2X$ (une seule observation doublée, variance $4V$) et $X_1 + X_2$ (deux observations indépendantes, variance $2V$).
- Additionne des variances, jamais des écarts types : $\sigma(S_n) = \sigma\sqrt{n}$ et non $n\sigma$.
- Pour appliquer Bienaymé-Tchebychev, écris l'événement sous la forme $|X - \mu| \geq \delta$, en vérifiant que l'intervalle est centré sur μ .
- Pour une taille d'échantillon, résous $V/(n\delta^2) \leq \alpha$; si p est inconnu, majore $p(1 - p)$ par $1/4$. Le n trouvé est suffisant, pas minimal.
- Une majoration supérieure ou égale à 1 n'apprend rien : c'est le signe que δ est trop petit devant σ .
- La loi des grands nombres porte sur la moyenne d'un grand nombre d'observations, jamais sur le résultat suivant.

Vers le PASS

En PASS et en LAS, ce chapitre fonde l'échantillonnage et l'estimation. La moyenne d'un échantillon de patients estime la moyenne de la population, avec une précision mesurée par l'écart type $\sigma/\sqrt{n}$, que l'on appelle **erreur standard** (ou écart type de la moyenne). L'additivité des variances sert à propager les incertitudes de mesure et à comparer deux moyennes (variance de la différence de deux moyennes indépendantes). Les intervalles de confiance et les calculs du nombre de sujets nécessaires dans un essai reposent sur la même logique que l'inégalité de concentration, avec des approximations plus fines par la loi normale. La loi des grands nombres justifie enfin la lecture d'une probabilité comme fréquence à long terme, notion constante en épidémiologie.

Suite en PASS : Échantillonnage et estimation par intervalle

Exercices

Niveau 1 : s'entraîner

1. X et Y sont deux variables aléatoires indépendantes avec $E(X) = 3$, $V(X) = 2$, $E(Y) = -1$ et $V(Y) = 5$. Calculer $E(2X - 3Y + 1)$, $V(2X - 3Y + 1)$ et l'écart type correspondant.
2. On considère un échantillon de taille 25 d'une loi d'espérance 10 et d'écart type 3. Calculer l'espérance, la variance et l'écart type de la somme S_{25} , puis de la moyenne M_{25} .
3. Une variable X a pour espérance 20 et pour variance 4.
- a) Minorer $P(16 < X < 24)$ à l'aide de l'inégalité de Bienaymé-Tchebychev. b) Majorer $P(|X - 20| \geq 10)$.
4. X suit la loi $B(40 ; 0,3)$. En écrivant X comme une somme de variables de Bernoulli, retrouver $E(X)$ et $V(X)$.
5. Une variable X a pour variance 3. On note X_1 et X_2 deux variables indépendantes de même loi que X . Calculer $V(2X)$ et $V(X_1 + X_2)$, puis expliquer la différence.

Niveau 2 : type bac

6. À chaque partie d'un jeu, le gain G d'un joueur vaut 2 € avec la probabilité 0,3 et -1 € avec la probabilité 0,7. Les parties sont indépendantes.
- a) Calculer $E(G)$ et $V(G)$. b) Le joueur joue 100 parties. On note S son gain total. Calculer $E(S)$, $V(S)$ et $\sigma(S)$. c) Le joueur joue 10 000 parties. Montrer, à l'aide de l'inégalité de Bienaymé-Tchebychev, que la probabilité que son gain total soit positif ou nul est inférieure à 0,02.
7. On interroge n personnes au hasard pour estimer la proportion p de personnes qui pratiquent un sport. On note F_n la fréquence observée.
- a) Donner $E(F_n)$ et montrer que $V(F_n) \leq 1/(4n)$. b) Pour $n = 1\,000$, majorer $P(|F_n - p| \geq 0,05)$. c) Déterminer une taille n suffisante pour que $P(|F_n - p| \geq 0,02) \leq 0,05$.
8. X suit la loi uniforme sur $\{0 ; 1 ; 2\}$ et Y la loi de Bernoulli de paramètre 0,5, X et Y étant indépendantes. On pose $S = X + Y$.
- a) Déterminer la loi de S . b) Calculer $E(S)$ de deux façons. c) Calculer $V(X)$, $V(Y)$, puis $V(S)$ de deux façons.
9. Une balance de précision commet sur chaque pesée une erreur aléatoire d'espérance nulle et d'écart type 0,4 mg, les erreurs des pesées successives étant indépendantes. On pèse n fois le même objet de masse m et l'on note M_n la moyenne des résultats.
- a) Exprimer $E(M_n)$ et $\sigma(M_n)$ en fonction de m et n . b) Combien de pesées faut-il au minimum pour que $\sigma(M_n) \leq 0,1$ mg ? c) Déterminer une valeur de n suffisante pour que $P(|M_n - m| \geq 0,2) \leq 0,05$.

Niveau 3 : vers le PASS

10. Démonstration par l'inégalité de Markov.
- a) Soit Y une variable aléatoire à valeurs positives ou nulles et a un réel strictement positif. Démontrer que $P(Y \geq a) \leq E(Y)/a$ (inégalité de Markov). b) En appliquant ce résultat à $Y = (X - \mu)^2$ et $a = \delta^2$, retrouver l'inégalité de Bienaymé-Tchebychev.
11. Soit X une variable de Bernoulli de paramètre $1/2$ et $Y = 1 - X$.

a) Calculer $V(X)$, $V(Y)$ et $V(X + Y)$. b) A-t-on $V(X + Y) = V(X) + V(Y)$? Expliquer pourquoi, en étudiant l'indépendance de X et Y .

12. La masse de principe actif d'un comprimé est une variable aléatoire d'espérance 500 mg et d'écart type 10 mg. Une boîte contient 30 comprimés, dont les masses sont indépendantes. On note S la masse totale de principe actif d'une boîte et M la masse moyenne par comprimé.

a) Calculer $E(S)$, $V(S)$ et $\sigma(S)$ (arrondi au dixième). b) Majorer la probabilité que S s'écarte de son espérance d'au moins 200 mg. c) Calculer $\sigma(M)$ à 10^{-2} près. Comparer à l'écart type d'un comprimé.

13. Vrai ou faux ? Justifier chaque réponse.

a) Si X et Y sont indépendantes, $V(X - Y) = V(X) - V(Y)$. b) L'égalité $E(X + Y) = E(X) + E(Y)$ exige que X et Y soient indépendantes. c) Après 10 piles consécutifs avec une pièce équilibrée, la probabilité d'obtenir face au lancer suivant est supérieure à $1/2$. d) Pour un échantillon de taille n d'une loi d'écart type σ , l'écart type de la moyenne vaut $\sigma/\sqrt{n}$. e) Pour toute variable X d'espérance μ et d'écart type $\sigma > 0$, $P(|X - \mu| < 2\sigma) \geq 0,75$.

14. Méthode de Monte-Carlo. On choisit au hasard un point dans le carré $[0 ; 1] \times [0 ; 1]$; la probabilité qu'il tombe dans le quart de disque de centre O et de rayon 1 vaut $p = \pi/4$. On répète n fois l'expérience de façon indépendante et l'on note F_n la fréquence des points tombés dans le quart de disque.

a) Quelle est la loi de nF_n ? Donner $E(F_n)$ et $V(F_n)$. b) Déterminer une valeur de n suffisante pour que $P(|F_n - \pi/4| \geq 0,01) \leq 0,1$, en utilisant la majoration $p(1 - p) \leq 1/4$. c) Pourquoi $4F_n$ fournit-il une estimation de π ? Quelle précision sur π l'écart de 0,01 sur F_n représente-t-il ?

Corrigés

1. Par linéarité : $E(2X - 3Y + 1) = 2E(X) - 3E(Y) + 1 = 6 + 3 + 1 = 10$.

X et Y étant indépendantes, $2X$ et $-3Y$ le sont aussi, et la constante ne modifie pas la variance : $V(2X - 3Y + 1) = 2^2V(X) + (-3)^2V(Y) = 4 \times 2 + 9 \times 5 = 8 + 45 = 53$. Écart type : $\sqrt{53} \approx 7,28$.

2. Avec $\mu = 10$, $V = 3^2 = 9$ et $n = 25$:

$E(S_{25}) = 25 \times 10 = 250$; $V(S_{25}) = 25 \times 9 = 225$; $\sigma(S_{25}) = \sqrt{225} = 15$. $E(M_{25}) = 10$; $V(M_{25}) = 9/25 = 0,36$; $\sigma(M_{25}) = 3/\sqrt{25} = 0,6$.

3. a) $16 < X < 24$ équivaut à $|X - 20| < 4$. Par Bienaymé-Tchebychev avec $\delta = 4$: $P(|X - 20| \geq 4) \leq 4/16 = 0,25$, donc $P(16 < X < 24) \geq 0,75$.

b) Avec $\delta = 10$: $P(|X - 20| \geq 10) \leq 4/100 = 0,04$.

4. $X = X_1 + \dots + X_{40}$, où X_i vaut 1 si la i -ème épreuve est un succès, 0 sinon ; les X_i sont indépendantes, de loi de Bernoulli de paramètre 0,3, avec $E(X_i) = 0,3$ et $V(X_i) = 0,3 \times 0,7 = 0,21$. Par linéarité : $E(X) = 40 \times 0,3 = 12$. Par indépendance : $V(X) = 40 \times 0,21 = 8,4$.

5. $V(2X) = 2^2V(X) = 12$. X_1 et X_2 étant indépendantes : $V(X_1 + X_2) = 3 + 3 = 6$. Explication : $2X$ double un seul résultat, donc double chaque écart à la moyenne et quadruple la variance ; $X_1 + X_2$ additionne deux résultats indépendants dont les écarts se compensent en partie (l'un peut être au-dessus de la moyenne, l'autre en dessous), d'où une variance seulement doublée.

6. a) $E(G) = 2 \times 0,3 + (-1) \times 0,7 = 0,6 - 0,7 = -0,1$ €. $E(G^2) = 4 \times 0,3 + 1 \times 0,7 = 1,9$, donc $V(G) = 1,9 - (-0,1)^2 = 1,9 - 0,01 = 1,89$.

b) S est la somme de 100 copies indépendantes de G : $E(S) = 100 \times (-0,1) = -10 \text{ €}$; $V(S) = 100 \times 1,89 = 189$; $\sigma(S) = \sqrt{189} \approx 13,75 \text{ €}$. c) Pour 10 000 parties : $E(S) = -1\,000$ et $V(S) = 18\,900$. Si $S \geq 0$, alors $S - (-1\,000) \geq 1\,000$, donc $|S + 1\,000| \geq 1\,000$. Ainsi $P(S \geq 0) \leq P(|S + 1\,000| \geq 1\,000) \leq 18\,900/1\,000^2 = 18\,900/1\,000\,000 = 0,0189$. Donc $P(S \geq 0) \leq 0,0189 < 0,02$. Sur un grand nombre de parties, le joueur est presque sûr de perdre.

7. a) nF_n , nombre de personnes sportives, suit $B(n, p)$; F_n est la moyenne d'un échantillon de la loi de Bernoulli de paramètre p . Donc $E(F_n) = p$ et $V(F_n) = p(1 - p)/n$. Or $p(1 - p) = 1/4 - (p - 1/2)^2 \leq 1/4$, d'où $V(F_n) \leq 1/(4n)$.

b) $P(|F_{1000} - p| \geq 0,05) \leq V(F_{1000})/0,05^2 \leq (1/4\,000)/0,0025 = 1/10 = 0,1$. c) $P(|F_n - p| \geq 0,02) \leq 1/(4n \times 0,0004) = 1/(0,0016n)$. Il suffit que $1/(0,0016n) \leq 0,05$, soit $n \geq 1/(0,0016 \times 0,05) = 1/0,00008 = 12\,500$ personnes.

8. a) $P(X = x) = 1/3$ pour $x \in \{0 ; 1 ; 2\}$, $P(Y = 0) = P(Y = 1) = 1/2$; par indépendance, chacun des 6 couples a la probabilité $1/6$.

$S = 0$: $(0 ; 0)$, probabilité $1/6$. $S = 1$: $(1 ; 0)$ ou $(0 ; 1)$, probabilité $2/6 = 1/3$. $S = 2$: $(2 ; 0)$ ou $(1 ; 1)$, probabilité $1/3$. $S = 3$: $(2 ; 1)$, probabilité $1/6$. **Loi de S** : $P(S = 0) = 1/6$, $P(S = 1) = 1/3$, $P(S = 2) = 1/3$, $P(S = 3) = 1/6$. b) Avec la loi : $E(S) = 0 \times 1/6 + 1 \times 1/3 + 2 \times 1/3 + 3 \times 1/6 = 1/3 + 2/3 + 1/2 = 1,5$. Par linéarité : $E(X) = 1$ et $E(Y) = 0,5$, donc $E(S) = 1,5$. c) $E(X^2) = (0 + 1 + 4)/3 = 5/3$, donc $V(X) = 5/3 - 1 = 2/3$; $V(Y) = 0,5 \times 0,5 = 1/4$. Par indépendance : $V(S) = 2/3 + 1/4 = 8/12 + 3/12 = 11/12$. Avec la loi : $E(S^2) = 1 \times 1/3 + 4 \times 1/3 + 9 \times 1/6 = 5/3 + 3/2 = 19/6$, et $V(S) = 19/6 - 1,5^2 = 38/12 - 27/12 = 11/12$. Les deux méthodes concordent.

9. a) Chaque résultat de pesée s'écrit $m + \varepsilon$, où l'erreur ε a pour espérance 0 et écart type 0,4 mg. Les résultats forment un échantillon d'espérance m et d'écart type 0,4. Donc $E(M_n) = m$ et $\sigma(M_n) = 0,4/\sqrt{n}$.

b) $0,4/\sqrt{n} \leq 0,1$ équivaut à $\sqrt{n} \geq 4$, soit $n \geq 16$: il faut au moins 16 pesées. c) $V(M_n) = 0,16/n$. Par l'inégalité de concentration : $P(|M_n - m| \geq 0,2) \leq (0,16/n)/0,04 = 4/n$. Il suffit que $4/n \leq 0,05$, soit $n \geq 80$ pesées.

10. a) Notons $y_1, \dots, y_p$ les valeurs (positives ou nulles) de Y , de probabilités $q_1, \dots, q_p$. $E(Y)$ est la somme des $q_i y_i$, tous positifs ou nuls ; en ne gardant que les indices tels que $y_i \geq a$, on obtient $E(Y) \geq$ somme des $q_i y_i$ pour $y_i \geq a \geq a \times$ (somme des q_i pour $y_i \geq a$) $= a \times P(Y \geq a)$. En divisant par $a > 0$: $P(Y \geq a) \leq E(Y)/a$.

b) $Y = (X - \mu)^2$ est positive, d'espérance $E((X - \mu)^2) = V(X)$. Avec $a = \delta^2 > 0$: $P((X - \mu)^2 \geq \delta^2) \leq V(X)/\delta^2$. Or, δ étant positif, $(X - \mu)^2 \geq \delta^2$ équivaut à $|X - \mu| \geq \delta$. On obtient $P(|X - \mu| \geq \delta) \leq V(X)/\delta^2$.

11. a) $V(X) = 1/2 \times 1/2 = 1/4$. $Y = 1 - X$ donne $V(Y) = (-1)^2 V(X) = 1/4$. Enfin $X + Y = 1$ est constante, donc $V(X + Y) = 0$.

b) **Non** : $V(X) + V(Y) = 1/2 \neq 0$. La formule exige l'indépendance, or X et Y ne sont pas indépendantes : $P(X = 1 \text{ et } Y = 1) = 0$, alors que $P(X = 1) \times P(Y = 1) = 1/2 \times 1/2 = 1/4$. Connaître X détermine entièrement Y .

12. a) $E(S) = 30 \times 500 = 15\,000 \text{ mg}$; $V(S) = 30 \times 10^2 = 3\,000 \text{ mg}^2$; $\sigma(S) = \sqrt{3\,000} \approx 54,8 \text{ mg}$.

b) Par Bienaymé-Tchebychev avec $\delta = 200$: $P(|S - 15\,000| \geq 200) \leq 3\,000/200^2 = 3\,000/40\,000 = 0,075$. c) $\sigma(M) = 10/\sqrt{30} \approx 1,83 \text{ mg}$, soit environ 5,5 fois moins que l'écart type d'un comprimé (10 mg) : la moyenne sur 30 comprimés est bien plus stable qu'un comprimé isolé.

13. a) **Faux** : pour X et Y indépendantes, $V(X - Y) = V(X) + (-1)^2 V(Y) = V(X) + V(Y)$.

b) **Faux** : la linéarité de l'espérance est vraie pour toutes variables, indépendantes ou non. c) **Faux** : les lancers sont indépendants, la probabilité d'obtenir face reste $1/2$. La loi des grands nombres ne prévoit aucune compensation. d) **Vrai** : $V(M_n) = \sigma^2/n$ par indépendance, donc $\sigma(M_n) = \sigma/\sqrt{n}$. e) **Vrai** : par Bienaymé-Tchebychev avec $\delta = 2\sigma$, $P(|X - \mu| \geq 2\sigma) \leq \sigma^2/(4\sigma^2) = 1/4$, donc $P(|X - \mu| < 2\sigma) \geq 3/4 = 0,75$.

14. a) Les n expériences sont des épreuves de Bernoulli identiques et indépendantes de succès « le point est dans le quart de disque », de probabilité $p = \pi/4$: nF_n suit $B(n ; \pi/4)$. Donc $E(F_n) = \pi/4$ et $V(F_n) = (\pi/4)(1 - \pi/4)/n$.

b) Comme $p(1 - p) \leq 1/4$, $V(F_n) \leq 1/(4n)$. Par l'inégalité de concentration : $P(|F_n - \pi/4| \geq 0,01) \leq 1/(4n \times 0,0001) = 1/(0,0004n)$. Il suffit que $1/(0,0004n) \leq 0,1$, soit $n \geq 1/0,00004 = 25\,000$ points. c) Par la loi des grands nombres, F_n se rapproche de $\pi/4$ quand n augmente, donc $4F_n$ se rapproche de π . Un écart de 0,01 sur F_n correspond à un écart de $4 \times 0,01 = 0,04$ sur l'estimation de π . Avec 25 000 points, l'estimation de π est donc à moins de 0,04 de π avec une probabilité d'au moins 0,9.