MATHÉMATIQUES

Maîtriser les sommes de variables aléatoires

Une fiche d'exercices pour additionner espérances et variances, raisonner sur un échantillon et utiliser les inégalités de concentration

Fiche de cours : Sommes de variables aléatoires.

① Pourquoi cette fiche ?

Ce chapitre repose sur quelques formules courtes, mais chacune a une condition ou un piège : l'espérance s'additionne toujours, la variance seulement avec indépendance, l'écart-type jamais ; la moyenne d'un échantillon se resserre en $\frac{\sigma}{\sqrt{n}}$ et non en $\frac{\sigma}{n}$; les inégalités donnent des majorants et non des valeurs.

Cette fiche isole chaque geste : calculer l'espérance et la variance d'une combinaison, additionner des variances à bon droit, démasquer les erreurs fréquentes, manier les formules de l'échantillon, appliquer et interpréter les inégalités de concentration, puis réinvestir tout cela dans deux problèmes où la loi des grands nombres explique une pratique réelle.

② Espérance et variance d'une combinaison

Rappel : ce qui s'additionne et ce qui ne s'additionne pas

$\mathrm{E}(a\mathrm{X}+b\mathrm{Y}) = a\mathrm{E}(\mathrm{X})+b\mathrm{E}(\mathrm{Y})$ toujours. $\mathrm{V}(a\mathrm{X}+b) = a^2\mathrm{V}(\mathrm{X})$ toujours. $\mathrm{V}(\mathrm{X}+\mathrm{Y}) = \mathrm{V}(\mathrm{X})+\mathrm{V}(\mathrm{Y})$ seulement si X et Y sont indépendantes. Les écarts-types ne s'additionnent jamais : on passe par les variances.

Exercice 1 *Transformations affines et sommes* La variable X prend les valeurs -2, 1 et 4 avec les probabilités 0,5, 0,3 et 0,2.

a) Calculer $\mathrm{E}(\mathrm{X})$ et $\mathrm{V}(\mathrm{X})$.

b) Soit $\mathrm{Y} = 3\mathrm{X}+2$. Calculer $\mathrm{E}(\mathrm{Y})$, $\mathrm{V}(\mathrm{Y})$ et comparer $\sigma(\mathrm{Y})$ à $\sigma(\mathrm{X})$.

c) Soit Z une variable indépendante de X avec $\mathrm{E}(\mathrm{Z}) = 4$ et $\mathrm{V}(\mathrm{Z}) = 1$. Calculer $\mathrm{E}(\mathrm{X}+\mathrm{Z})$ et $\mathrm{V}(\mathrm{X}+\mathrm{Z})$.

d) Un professeur ajoute 2 points à toutes les notes d'un devoir, puis un autre les multiplie par 1,1. Que deviennent la moyenne et l'écart-type dans chaque cas ?

③ Indépendance et variance

Exercice 2 *Quand peut-on additionner les variances ?*

a) On lance deux dés ; X et Y sont les résultats. Vérifier que $P(X = 6 \text{ et } Y = 6) = P(X = 6) \times P(Y = 6)$. Que traduit cette égalité ?

b) X est le résultat d'un dé et $Z = 7 - X$. Calculer $E(Z)$ et $V(Z)$, puis $V(X + Z)$. Comparer avec $V(X) + V(Z)$ et expliquer.

c) X suit $\mathcal{B}(4\ ;\ 0{,}5)$ et Y suit $\mathcal{B}(6\ ;\ 0{,}5)$, indépendantes. Loi, espérance et variance de $X + Y$?

d) Trois variables indépendantes ont chacune un écart-type de 2. Quel est l'écart-type de leur somme ?

④ Démasquer les erreurs

Exercice 3 *Cinq raisonnements à corriger* Chaque raisonnement contient une erreur. La repérer et la corriger.

a) « X et Y indépendantes d'écarts-types 3 et 4 : $\sigma(X + Y) = 7$. »

b) « X et Y indépendantes : $V(X - Y) = V(X) - V(Y)$. »

c) « Pour un échantillon de taille n, $\sigma(M_n) = \dfrac{\sigma}{n}$. »

d) « D'après la loi des grands nombres, sur 1 000 lancers d'une pièce on obtiendra exactement 500 piles. »

e) « $E(X) = 20$, $V(X) = 9$, donc $P(|X - 20| \geqslant 6) = \dfrac{9}{36} = 0{,}25$. »

⑤ Échantillons

Rappel : somme ou moyenne ?

Pour n répétitions indépendantes : $E(S_n) = n\mu$ et $\sigma(S_n) = \sqrt{n}\,\sigma$ (la somme se disperse davantage) ; $E(M_n) = \mu$ et $\sigma(M_n) = \dfrac{\sigma}{\sqrt{n}}$ (la moyenne se resserre). Bien lire si la question porte sur un total ou sur une moyenne.

Exercice 4 *Notes, ampoules et effectifs*

a) La note X à un devoir, dans une grande population, a pour espérance 11 et pour écart-type 3. On considère une classe de 36 élèves comme un échantillon. Calculer $E(M_{36})$ et $\sigma(M_{36})$.

b) Quelle taille d'échantillon donnerait $\sigma(M_n) \leqslant 0{,}25$?

c) Calculer l'espérance et l'écart-type de la somme S_{36} des notes de la classe.

d) La durée de vie d'une ampoule a pour espérance 1 000 h et pour écart-type 100 h. Pour un lot de 400 ampoules, donner l'espérance et l'écart-type de la durée totale et de la durée moyenne.

6 Inégalités de concentration

Exercice 5 *Majorer, minorer, comparer* Soit X une variable d'espérance 50 et de variance 25.

a) Majorer $P(|X - 50| \geqslant 10)$, puis minorer $P(40 < X < 60)$.

b) Majorer $P(|X - 50| \geqslant 15)$.

c) On suppose maintenant que X suit $\mathcal{B}(100\ ;\ 0{,}5)$ (même espérance, même variance). La calculatrice donne $P(X \leqslant 40) \approx 0{,}028$. Calculer $P(|X - 50| \geqslant 10)$ et comparer au majorant de la question a).

d) Pourquoi l'inégalité de Bienaymé-Tchebychev est-elle utile malgré ce manque de précision ?

Exercice 6 *Dimensionner un sondage* Dans une population, une proportion inconnue p de personnes a une certaine opinion. On interroge n personnes au hasard (avec remise, ou dans une population très grande) et F_n est la fréquence observée.

a) Justifier que F_n est la moyenne d'un échantillon d'une variable de Bernoulli de paramètre p, et que sa variance vérifie $V(X) = p(1-p) \leqslant \dfrac{1}{4}$.

b) En déduire que $P(|F_n - p| \geqslant 0{,}03) \leqslant \dfrac{1}{0{,}0036\,n}$.

c) Que vaut ce majorant pour $n = 1\,000$? Quelle taille garantit un majorant de 0,05 ?

d) Les instituts de sondage interrogent souvent 1 000 personnes et annoncent une précision d'environ 3 points. Que peut-on en dire au vu de la question précédente ?

⑦ Problèmes en contexte

Exercice 7 *La charge d'un ascenseur* La masse d'une personne (en kg) a pour espérance 75 et pour écart-type 15. Un ascenseur admet une charge maximale de 900 kg.

a) Dix personnes montent. Donner l'espérance et l'écart-type de la masse totale S_{10}.

b) Montrer que « la charge dépasse 900 kg » entraîne $|S_{10} - 750| \geqslant 150$, et majorer sa probabilité.

c) Reprendre avec huit personnes. Commenter.

Exercice 8 *Pourquoi le casino gagne toujours* À un jeu de casino, le gain X d'un joueur sur une partie (en euros, négatif en cas de perte) a pour espérance $-0{,}05$ et pour écart-type 2. Dans une journée, n parties indépendantes sont jouées ; S_n est le gain total des joueurs (donc la perte du casino).

a) Pour $n = 10\,000$, calculer $E(S_n)$ et $\sigma(S_n)$.

b) Le casino perd de l'argent si $S_n \geqslant 0$. Montrer que cet événement entraîne $|S_n + 500| \geqslant 500$ et majorer sa probabilité.

c) Reprendre pour $n = 1\,000\,000$ parties. Conclure.

d) Un joueur isolé fait 10 parties. Que valent $E(S_{10})$ et $\sigma(S_{10})$? Pourquoi peut-il malgré tout repartir gagnant ?

⑧ Pour s'auto-évaluer

Rappel : cinq questions à se poser

- Avant d'additionner des variances, ai-je vérifié l'indépendance ?
- Ai-je additionné des variances et non des écarts-types ?
- La question porte-t-elle sur une somme ($\sqrt{n}\,\sigma$) ou sur une moyenne ($\sigma/\sqrt{n}$) ?
- Ai-je écrit l'événement sous la forme $|X - E(X)| \geqslant \delta$ avant d'appliquer l'inégalité ?
- Ai-je présenté le résultat comme un majorant, et non comme la valeur de la probabilité ?

Solutions des exercices

Corrigé de l'exercice 1.

a) $E(X) = -2 \times 0{,}5 + 1 \times 0{,}3 + 4 \times 0{,}2 = 0{,}1$. $E(X^2) = 4 \times 0{,}5 + 1 \times 0{,}3 + 16 \times 0{,}2 = 5{,}5$, donc $V(X) = 5{,}5 - 0{,}01 = 5{,}49$.

b) $E(Y) = 3 \times 0{,}1 + 2 = 2{,}3$; $V(Y) = 9 \times 5{,}49 = 49{,}41$; $\sigma(Y) = 3\sigma(X)$ (la constante 2 ne change pas la dispersion).

c) $E(X + Z) = 4{,}1$; par indépendance $V(X + Z) = 5{,}49 + 1 = 6{,}49$.

d) Ajouter 2 : moyenne +2, écart-type inchangé. Multiplier par 1,1 : moyenne et écart-type multipliés par 1,1 (la variance par 1,21).

Corrigé de l'exercice 2.

a) $P(X = 6 \text{ et } Y = 6) = \dfrac{1}{36} = \dfrac{1}{6} \times \dfrac{1}{6}$: les deux dés sont indépendants (le résultat de l'un n'influence pas l'autre).

b) $E(Z) = 7 - 3{,}5 = 3{,}5$ et $V(Z) = (-1)^2 V(X) = \dfrac{35}{12}$. Mais $X + Z = 7$ est constante : $V(X + Z) = 0 \neq \dfrac{35}{6}$. X et Z ne sont pas indépendantes (l'une détermine l'autre), la formule ne s'applique pas.

c) $X + Y$ compte les succès de 10 épreuves indépendantes : loi $\mathcal{B}(10\ ;\ 0{,}5)$, $E = 5$, $V = 2{,}5$. Cohérent avec $E(X) + E(Y) = 2 + 3$ et $V(X) + V(Y) = 1 + 1{,}5$. ✓

d) Variance de la somme : $4 + 4 + 4 = 12$, donc écart-type $\sqrt{12} = 2\sqrt{3} \approx 3{,}46$, et non 6.

Corrigé de l'exercice 3.

a) On additionne les variances : $V(X + Y) = 9 + 16 = 25$, donc $\sigma(X + Y) = 5$.

b) $V(-Y) = V(Y)$, donc $V(X - Y) = V(X) + V(Y)$. Une variance ne peut pas être négative.

c) C'est la variance qui est divisée par n : $V(M_n) = \dfrac{\sigma^2}{n}$ et $\sigma(M_n) = \dfrac{\sigma}{\sqrt{n}}$.

d) La loi des grands nombres dit que la fréquence sera **probablement proche** de 0,5, pas exactement égale : le nombre de piles reste aléatoire, d'écart-type $\sqrt{250} \approx 16$.

e) L'inégalité de Bienaymé-Tchebychev donne une majoration : $P(|X - 20| \geqslant 6) \leqslant 0{,}25$. La vraie valeur peut être bien plus petite.

Corrigé de l'exercice 4.

a) $E(M_{36}) = 11$ et $\sigma(M_{36}) = \dfrac{3}{\sqrt{36}} = 0{,}5$.

b) $\dfrac{3}{\sqrt{n}} \leqslant 0{,}25 \iff \sqrt{n} \geqslant 12 \iff n \geqslant 144$.

c) $E(S_{36}) = 36 \times 11 = 396$ et $\sigma(S_{36}) = \sqrt{36} \times 3 = 18$.

d) Durée totale : $E = 400\,000$ h, $\sigma = \sqrt{400} \times 100 = 2\,000$ h. Durée moyenne : $E = 1\,000$ h, $\sigma = \dfrac{100}{20} = 5$ h.

Corrigé de l'exercice 5.

a) $P(|X - 50| \geqslant 10) \leqslant \dfrac{25}{100} = 0{,}25$, donc $P(40 < X < 60) \geqslant 0{,}75$.

b) $P(|X - 50| \geqslant 15) \leqslant \dfrac{25}{225} = \dfrac{1}{9} \approx 0{,}11$.

c) Par symétrie de $\mathcal{B}(100\,;\,0{,}5)$: $P(|X-50| \geqslant 10) = P(X \leqslant 40) + P(X \geqslant 60) \approx 2 \times 0{,}028 = 0{,}056$, plus de quatre fois moins que le majorant 0,25.

d) Elle ne suppose rien sur la loi de X : seules l'espérance et la variance sont nécessaires. Elle s'applique donc même quand la loi est inconnue, ce qui est le cas de la plupart des situations réelles (sondages, mesures).

Corrigé de l'exercice 6.

a) Chaque personne interrogée donne une variable X_i valant 1 (opinion) avec la probabilité p, 0 sinon ; les X_i sont indépendantes et F_n est leur moyenne. $V(X) = p(1-p)$; la fonction $p \mapsto p(1-p)$ est une parabole tournée vers le bas, maximale en $p = \dfrac{1}{2}$ où elle vaut $\dfrac{1}{4}$.

b) Inégalité de concentration : $P(|F_n - p| \geqslant 0{,}03) \leqslant \dfrac{V(X)}{n \times 0{,}03^2} \leqslant \dfrac{1/4}{0{,}0009\,n} = \dfrac{1}{0{,}0036\,n}$.

c) $n = 1\,000$: $\dfrac{1}{3{,}6} \approx 0{,}28$. Pour 0,05 : $n \geqslant \dfrac{1}{0{,}0036 \times 0{,}05} \approx 5\,556$ personnes.

d) Avec les seuls outils de cette fiche, 1 000 personnes ne garantissent la précision de 3 points qu'avec une probabilité d'au moins 0,72. Les instituts utilisent des résultats plus fins sur la loi de F_n (hors programme), qui donnent pour $n = 1\,000$ une garantie d'environ 95 % : l'inégalité de concentration est vraie mais très prudente.

Corrigé de l'exercice 7.

a) $E(S_{10}) = 750$ kg et $\sigma(S_{10}) = \sqrt{10} \times 15 \approx 47{,}4$ kg (variance 2 250).

b) $S_{10} > 900 \iff S_{10} - 750 > 150$, qui entraîne $|S_{10} - 750| \geqslant 150$. Bienaymé-Tchebychev : $P \leqslant \dfrac{2\,250}{150^2} = \dfrac{2\,250}{22\,500} = 0{,}1$.

c) Huit personnes : $E = 600$, $V = 8 \times 225 = 1\,800$. Dépasser 900 impose $|S_8 - 600| \geqslant 300$: $P \leqslant \dfrac{1\,800}{90\,000} = 0{,}02$. Limiter à huit personnes divise le majorant par cinq ; la vraie probabilité est bien plus faible encore, l'inégalité étant prudente.

Corrigé de l'exercice 8.

a) $E(S_n) = 10\,000 \times (-0{,}05) = -500$ € et $\sigma(S_n) = \sqrt{10\,000} \times 2 = 200$ €.

b) $S_n \geqslant 0 \iff S_n - (-500) \geqslant 500$, qui entraîne $|S_n + 500| \geqslant 500$. Bienaymé-Tchebychev : $P \leqslant \dfrac{V(S_n)}{500^2} = \dfrac{40\,000}{250\,000} = 0{,}16$.

c) $E(S_n) = -50\,000$ €, $V(S_n) = 4\times10^6$, $\sigma(S_n) = 2\,000$ €. Perte du casino si $|S_n + 50\,000| \geqslant 50\,000$: $P \leqslant \dfrac{4\times10^6}{2{,}5\times10^9} = 0{,}0016$. Sur un grand volume de parties, le casino gagne avec une probabilité proche de 1 : c'est la loi des grands nombres qui travaille pour lui.

d) $E(S_{10}) = -0{,}5$ € et $\sigma(S_{10}) = \sqrt{10}\times 2 \approx 6{,}3$ €. L'écart-type est bien plus grand que l'espérance en valeur absolue : sur peu de parties, le hasard domine et un gain positif est tout à fait possible. C'est sur le grand nombre que l'avantage du casino devient une quasi-certitude.