Chapitre 13. Cours

Sommes de variables aléatoires et échantillonnage

1. Rappels sur une variable aléatoire

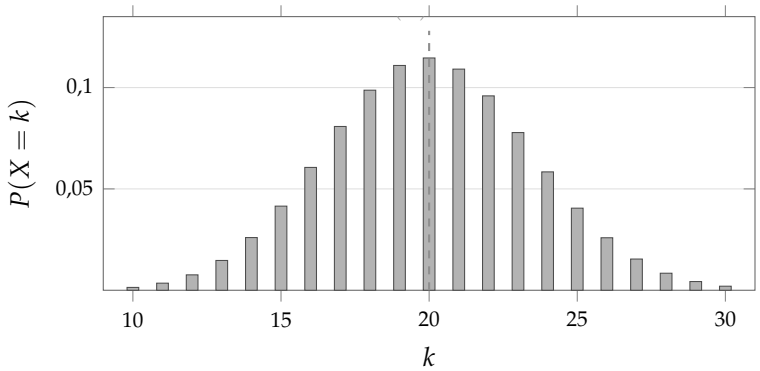


Figure 1 — Loi $\mathcal{B}(50\,;0{,}4)$: la distribution se concentre autour de son espérance.

Propriété 13.1 — Espérance, variance, écart-type

Si X prend les valeurs $x_1, \dots, x_k$ avec les probabilités $p_1, \dots, p_k$:

$$E(X) = \sum_{i=1}^{k} p_i\, x_i \qquad V(X) = E\left(X^2\right) - \left(E(X)\right)^2 \qquad \sigma(X) = \sqrt{V(X)}.$$

Exercice d'application 1

X est le résultat d'un dé équilibré. Calculer $E(X)$, $V(X)$ et $\sigma(X)$.

Solution. $E(X) = \dfrac{1+2+3+4+5+6}{6} = \dfrac{7}{2} = 3{,}5.$

$E\left(X^2\right) = \dfrac{1+4+9+16+25+36}{6} = \dfrac{91}{6}$, donc

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182-147}{12} = \frac{35}{12} \qquad \text{et} \qquad \sigma(X) = \sqrt{\frac{35}{12}} \approx 1{,}708.$$

Remarque — Bon réflexe

La formule $V(X) = E\left(X^2\right) - \mu^2$ est presque toujours plus rapide que la définition par les écarts. Calculez $E\left(X^2\right)$ en élevant les **valeurs** au carré, jamais les probabilités.

2. Transformation affine

Théorème 13.2 — Transformation affine

Pour tous réels a et b :

$$E(aX+b) = a\,E(X) + b \qquad V(aX+b) = a^2\,V(X) \qquad \sigma(aX+b) = |a|\,\sigma(X).$$

Lecture. Le décalage b translate la moyenne sans changer la dispersion ; le facteur a dilate les deux, mais la variance suit le **carré**.

Exercice d'application 2

On sait que $E(X) = 4$ ***et*** $\sigma(X) = 2$***. Calculer*** $E(Y)$ ***et*** $\sigma(Y)$ ***pour*** $Y = -3X + 5$**.**

Solution. $E(Y) = -3 \times 4 + 5 = -7$, et $\sigma(Y) = |-3| \times 2 = 6$.

Attention — Erreur à éviter

Écrire $V(aX+b) = a\,V(X) + b$. La variance ignore b et récupère a^2 : c'est la conséquence directe de sa définition par des carrés d'écarts.

3. Somme de deux variables aléatoires

Théorème 13.3 — Linéarité de l'espérance

Pour toutes variables X et Y :

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

Théorème 13.4 — Variance d'une somme

Si X et Y sont **indépendantes** :

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

Exercice d'application 3

On lance un dé (X***, application 1) et une pièce (*** Y ***vaut 1 pour pile, 0 pour face), indépendamment. Calculer*** $E(X+Y)$ ***et*** $V(X+Y)$**.**

Solution. $E(Y) = \dfrac{1}{2}$ et $V(Y) = E\left(Y^2\right) - \dfrac{1}{4} = \dfrac{1}{2} - \dfrac{1}{4} = \dfrac{1}{4}$.

Par linéarité : $E(X+Y) = \dfrac{7}{2} + \dfrac{1}{2} = 4$.

Par indépendance : $V(X+Y) = \dfrac{35}{12} + \dfrac{1}{4} = \dfrac{38}{12} = \dfrac{19}{6}$.

Attention — Erreur à éviter

L'espérance est linéaire **sans condition** ; la variance n'est additive que sous l'hypothèse d'**indépendance**. Oublier de la vérifier est l'erreur la plus coûteuse du chapitre.

Remarque — Astuce

Les écarts-types, eux, ne s'additionnent **jamais** : $\sigma(X+Y) = \sqrt{V(X)+V(Y)}$, ce qui est strictement inférieur à $\sigma(X)+\sigma(Y)$. Les dispersions se compensent partiellement.

4. Application à la loi binomiale

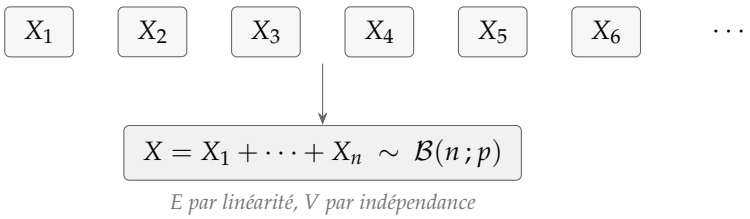


Figure 2 — La binomiale, somme de n variables de Bernoulli.

Propriété 13.5 — La binomiale est une somme

Une variable X suivant $\mathcal{B}(n\,;p)$ s'écrit comme la somme de n variables de Bernoulli **indépendantes** $X_1, \ldots, X_n$ de paramètre p.

Démonstration — des formules du chapitre 7

Chaque X_i vérifie $E(X_i) = p$ et $V(X_i) = p(1-p)$. Par linéarité et additivité :

$$E(X) = np \qquad V(X) = np(1-p). \qquad \blacksquare$$

Exercice d'application 4

Retrouver l'espérance et la variance de la loi $\mathcal{B}(50\,;0{,}4)$ de la figure 1.

Solution. X est la somme de 50 variables de Bernoulli indépendantes de paramètre 0,4 :

$$E(X) = 50 \times 0{,}4 = 20 \qquad \text{et} \qquad V(X) = 50 \times 0{,}4 \times 0{,}6 = 12.$$

Remarque — Ce qu'il faut voir

Les formules du chapitre 7, admises alors, se **démontrent** ici en trois lignes. C'est une démonstration exigible, et la décomposition en somme est l'outil qui la rend possible.

5. Échantillon

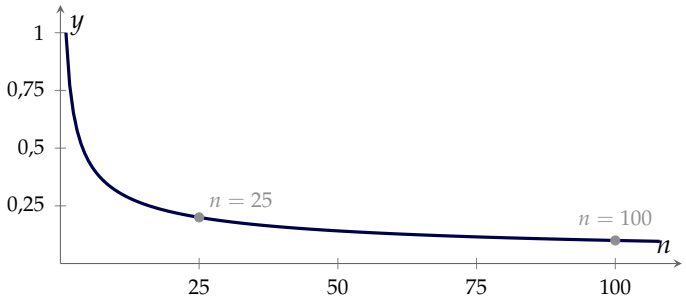


Figure 3 — L'écart-type de la moyenne décroît en $\dfrac{\sigma}{\sqrt{n}}$.

Définition 13.6 — Échantillon

Un **échantillon** de taille n est une suite $X_1, \dots, X_n$ de variables aléatoires **indépendantes et de même loi**, d'espérance μ et de variance V.
On pose $S_n = X_1 + \cdots + X_n$ et $M_n = \dfrac{S_n}{n}$, la **moyenne empirique**.

Théorème 13.7 — Somme et moyenne d'un échantillon

$$E(S_n) = n\mu \qquad V(S_n) = nV \qquad \sigma(S_n) = \sqrt{n}\,\sigma$$
$$E(M_n) = \mu \qquad V(M_n) = \frac{V}{n} \qquad \sigma(M_n) = \frac{\sigma}{\sqrt{n}}.$$

Exercice d'application 5

On lance 100 ***fois le dé de l'application 1 et l'on note*** M_{100} ***la moyenne obtenue. Calculer*** $E(M_{100})$ ***et*** $\sigma(M_{100})$**.**

Solution. Les lancers forment un échantillon de taille 100, avec $\mu = 3{,}5$ et $\sigma = \sqrt{\dfrac{35}{12}}$:

$$E(M_{100}) = 3{,}5 \qquad \text{et} \qquad \sigma(M_{100}) = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10} \approx 0{,}171.$$

Remarque — Le résultat central

La moyenne empirique a la **même espérance** que chaque terme, mais un écart-type **divisé par** $\sqrt{n}$. C'est la racine, pas n lui-même : diviser la dispersion par deux exige de **quadrupler** l'échantillon. Tout le chapitre 14 découle de cette formule.

Exercice d'application 6

Quelle taille d'échantillon garantit $\sigma(M_n) \leqslant 0{,}05$ ***pour le dé de l'application 1 ?***

Solution. La contrainte s'écrit $\dfrac{\sigma}{\sqrt{n}} \leqslant 0{,}05$, soit $\sqrt{n} \geqslant \dfrac{\sigma}{0{,}05}$ puis, en élevant au carré :

$$n \geqslant \frac{35/12}{0{,}05^2} = \frac{35}{12 \times 0{,}0025} \approx 1166{,}7.$$

Le plus petit entier convenable est $n = 1167$.
Vérification : $\dfrac{1{,}70783}{\sqrt{1167}} \approx 0{,}04999 \leqslant 0{,}05$ ✓.

Attention — Erreur à éviter

Confondre S_n et M_n. La somme se disperse de plus en plus — son écart-type croît comme $\sqrt{n}$; la moyenne se concentre — le sien décroît comme $\dfrac{1}{\sqrt{n}}$.

Méthodes

Méthode 1 — Espérance et variance d'une transformation affine

1. Identifier a et b dans l'écriture $Y = aX + b$.
2. Appliquer $E(Y) = aE(X) + b$.
3. Appliquer $V(Y) = a^2V(X)$ — le b disparaît.
4. Pour l'écart-type, prendre $|a|\,\sigma(X)$, valeur absolue comprise.

Exemple éclair. Si $E(X) = 4$ et $\sigma(X) = 2$, alors $Y = -3X + 5$ vérifie $E(Y) = -7$ et $\sigma(Y) = 6$.
Erreur à éviter. Omettre la valeur absolue. Un écart-type est toujours positif : $\sigma(-3X) = 3\,\sigma(X)$, jamais $-3\,\sigma(X)$.

Méthode 2 — Décomposer en somme de variables simples

1. Repérer que la variable compte un total ou un cumul.
2. Écrire $X = X_1 + \cdots + X_n$ en précisant ce que vaut chaque X_i.
3. Calculer $E(X_i)$ et $V(X_i)$ pour une seule d'entre elles.
4. Appliquer la linéarité, et l'additivité des variances **après avoir justifié l'indépendance**.

Exemple éclair. $X \sim \mathcal{B}(n\,;p)$: chaque X_i de Bernoulli donne $E(X) = np$ et $V(X) = np(1-p)$ en trois lignes.
Astuce. Cette décomposition transforme un calcul global en un calcul élémentaire répété, sans jamais manipuler de coefficient binomial.

Méthode 3 — Étudier un échantillon

1. Vérifier que les variables sont **indépendantes et de même loi**.
2. Relever μ et V pour une variable individuelle.
3. Appliquer les formules de S_n ou de M_n selon ce qui est demandé.
4. Interpréter : la somme se disperse, la moyenne se concentre.

Exemple éclair. 100 lancers du dé : $E(M_{100}) = 3{,}5$ et $\sigma(M_{100}) \approx 0{,}171$ — c'est l'application 5.
Erreur à éviter. Utiliser les formules de S_n quand la question porte sur M_n. Lisez l'énoncé : « total » appelle S_n, « moyenne » ou « proportion » appelle M_n.

Méthode 4 — Dimensionner un échantillon

1. Écrire la contrainte sur l'écart-type : $\dfrac{\sigma}{\sqrt{n}} \leqslant s$.
2. Isoler $\sqrt{n}$, puis élever au carré : $n \geqslant \dfrac{\sigma^2}{s^2}$.
3. Prendre le plus petit entier convenable.
4. Vérifier en recalculant $\dfrac{\sigma}{\sqrt{n}}$.

Exemple éclair. $\sigma(M_n) \leqslant 0{,}05$ pour le dé : $n \geqslant \dfrac{35/12}{0{,}0025} \approx 1166{,}7$, donc $n = 1167$ — c'est l'application 6.
Bon réflexe. L'exposant 2 est le cœur de la question : diviser la dispersion par k multiplie la taille par k^2. C'est la même mécanique que la marge d'erreur du chapitre 14.

Méthode 5 — Calculer l'espérance d'un gain

1. Dresser la loi de probabilité du gain, valeurs et probabilités.
2. Vérifier que la somme des probabilités vaut 1.
3. Calculer $E(G) = \sum p_i\, g_i$, **sans oublier** de compter la mise en négatif.
4. Conclure : jeu favorable si $E(G) > 0$, équitable si $E(G) = 0$, défavorable sinon.

Exemple éclair. Mise 2 euros, gain brut 5 euros avec probabilité $\dfrac{1}{3}$: $E(G) = \dfrac{1}{3} \times 3 + \dfrac{2}{3} \times (-2) = -\dfrac{1}{3}$ — jeu défavorable.
Erreur à éviter. Confondre le gain **brut** et le gain **algébrique**. Si la partie coûte 2 euros et rapporte 5, le gain algébrique est $+3$, pas $+5$.

Méthode 6 — Justifier une indépendance, ou son absence

Quand ? L'énoncé demande d'additionner des variances, ce qui suppose l'indépendance — ou propose une situation où elle échoue.
1. Décrire précisément l'expérience associée à chaque variable.
2. Se demander si le résultat de l'une modifie la loi de l'autre.
3. Tirage avec remise, lancers répétés, individus distincts d'une grande population : indépendance.
4. Tirage sans remise, variables portant sur le même individu : **pas** d'indépendance — la variance ne s'additionne plus.

Exemple éclair. Dé et pièce lancés séparément : indépendance évidente, $V(X+Y) = V(X) + V(Y)$ — c'est l'application 3.
Bon réflexe. L'espérance reste linéaire même sans indépendance. Si vous n'avez besoin que de $E(X+Y)$, la question ne se pose pas — et le dire vaut un point.

Erreurs fréquentes et bons réflexes

Erreur	Bon réflexe
Écrire $V(aX+b) = aV(X) + b$	C'est $a^2V(X)$, et b disparaît
Additionner les variances sans indépendance	Le théorème 13.4 l'exige
Exiger l'indépendance pour l'espérance	La linéarité est inconditionnelle
Écrire $\sigma(M_n) = \dfrac{\sigma}{n}$	C'est $\dfrac{\sigma}{\sqrt{n}}$
Confondre V et σ	La variance est le carré de l'écart-type
Croire que doubler n divise l'incertitude par 2	Elle est divisée par $\sqrt{2} \approx 1{,}41$

Module Python — convergence de la moyenne d'un échantillon

```
# Convergence de la moyenne d'un echantillon
from random import randint
def moyenne_echantillon(n):
  # n lancers d'un de equilibre
  somme = 0
  for i in range(n):
     somme = somme + randint(1, 6)
  return somme / n
mu = 3.5
sigma = (35 / 12) ** 0.5
print("Loi du de :  mu =", mu, " sigma =", round(sigma, 6))
liste_n = [10, 100, 1000, 10000, 100000]
for i in range(len(liste_n)):
  n = liste_n[i]
  m = moyenne_echantillon(n)
  ecart_type = sigma / n ** 0.5
  print("n =", n, " moyenne :", round(m, 4))
  print(" ecart-type de M_n :", round(ecart_type, 6))
```

Trace d'exécution *(les moyennes varient d'une exécution à l'autre).*

```
Loi du de :  mu = 3.5 sigma = 1.707825
n = 10 moyenne :  3.6
ecart-type de M_n :  0.540062
n = 100 moyenne :  3.35
ecart-type de M_n :  0.170783
n = 1000 moyenne :  3.471
ecart-type de M_n :  0.054006
n = 10000 moyenne :  3.5221
ecart-type de M_n :  0.017078
n = 100000 moyenne :  3.5012
ecart-type de M_n :  0.005401
```

Interprétation mathématique. L'écart-type théorique de M_n vaut $\dfrac{\sigma}{\sqrt{n}}$. Il est divisé par $\sqrt{10} \approx 3{,}16$ chaque fois que n est multiplié par 10 : on lit $0{,}540 \to 0{,}171 \to 0{,}054 \to 0{,}017 \to 0{,}0054$, rapports conformes. ✓

Les écarts observés entre la moyenne empirique et 3,5 restent, à chaque ligne, du même ordre de grandeur que $\dfrac{\sigma}{\sqrt{n}}$ — parfois un peu au-dessus, le plus souvent en dessous. Relancez le programme : les moyennes changent, la colonne des écarts-types théoriques, elle, ne bouge pas.

Attention

Ce n'est pas une démonstration. La trace montre que M_n s'approche de μ, mais rien n'interdit une exécution où $M_{100\,000}$ s'écarterait fortement — c'est simplement très improbable. **Quantifier ce « très improbable » est l'objet du chapitre 14.**