Chapitre 14. Cours

Concentration et loi des grands nombres

1. Inégalité de Markov

Le chapitre 13 a montré que la moyenne d'un échantillon se concentre. Ce chapitre **quantifie** cette concentration, par une chaîne de trois inégalités qui aboutit à la loi des grands nombres.

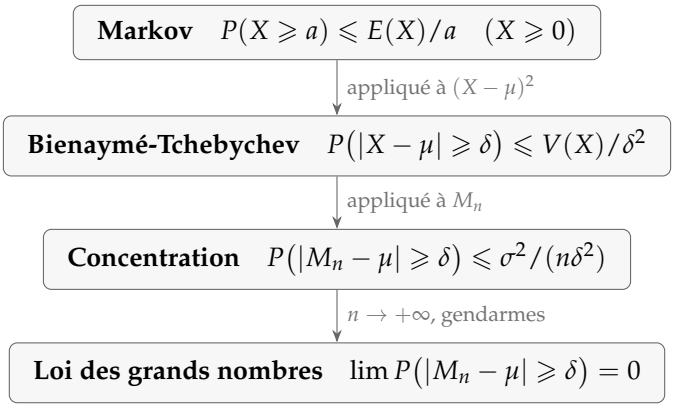


Figure 1 — De Markov à la loi des grands nombres.

Théorème 14.1 — Markov

(complément utile à la démonstration, non exigible) Soit X une variable aléatoire **à valeurs positives ou nulles**. Pour tout réel $a > 0$:

$$P(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Exercice d'application 1

Un site reçoit en moyenne 200 visites par heure. Majorer la probabilité d'en recevoir au moins 1 000 dans une heure donnée.

Solution. Le nombre de visites est positif, d'espérance 200. Markov donne :

$$P(X \geqslant 1\,000) \leqslant \frac{200}{1\,000} = 0{,}2.$$

Attention — Erreur à éviter

L'hypothèse $X \geqslant 0$ est **indispensable**. Sans elle, l'inégalité est fausse : une variable pouvant prendre de grandes valeurs négatives aurait une espérance faible tout en dépassant souvent a.

Remarque — Ce que dit Markov

Une seule information — l'espérance — suffit à majorer la probabilité des grandes valeurs. C'est une majoration **grossière**, mais elle ne demande rien d'autre.

2. Inégalité de Bienaymé-Tchebychev

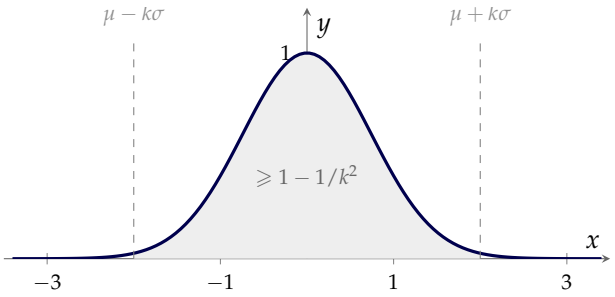


Figure 2 — La zone contrôlée par l'inégalité.

Théorème 14.2 — Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance μ et de variance V. Pour tout réel $\delta > 0$:

$$P\big(|X - \mu| \geqslant \delta\big) \leqslant \frac{V}{\delta^2}.$$

Lecture. L'écart à la moyenne est d'autant plus improbable que la variance est faible et l'écart exigé grand. Le carré au dénominateur fait chuter la majoration très vite.

Exercice d'application 2

*X **suit la loi** $\mathcal{B}(100\,;0{,}5)$. **Majorer** $P\big(|X - 50| \geqslant 20\big)$.*

Solution. Ici $\mu = np = 50$ et $V = np(1-p) = 25$. Bienaymé-Tchebychev donne :

$$P\big(|X - 50| \geqslant 20\big) \leqslant \frac{25}{20^2} = \frac{25}{400} = 0{,}0625.$$

Remarque — Bon réflexe

Contrairement à Markov, cette inégalité ne demande **aucune** hypothèse de signe. En revanche elle exige la variance, donc une information de plus.

Propriété 14.3 — Forme en écarts-types

En posant $\delta = k\sigma$:

$$P\big(|X - \mu| \geqslant k\sigma\big) \leqslant \frac{1}{k^2}.$$

Au-delà de deux écarts-types, la probabilité est majorée par 0,25 ; au-delà de trois, par $\dfrac{1}{9}$.

Exercice d'application 3

*X **a pour espérance** 20 **et pour écart-type** 4. **Majorer** $P\big(|X - 20| \geqslant 8\big)$.*

Solution. L'écart 8 vaut $k = 2$ écarts-types :

$$P\big(|X - 20| \geqslant 8\big) \leqslant \frac{1}{2^2} = 0{,}25.$$

3. Inégalité de concentration

Théorème 14.4 — Concentration

Soit $X_1, \dots, X_n$ des variables aléatoires **indépendantes** de même loi, d'espérance μ et de variance V. On note $M_n = \dfrac{X_1 + \cdots + X_n}{n}$ la moyenne empirique. Pour tout $\delta > 0$:

$$P\big(|M_n - \mu| \geqslant \delta\big) \leqslant \frac{V}{n\,\delta^2}.$$

Démonstration

On applique Bienaymé-Tchebychev à M_n, dont l'espérance vaut μ et la variance $\dfrac{V}{n}$ (théorème 13.7). ■

Exercice d'application 4

On lance 100 fois un dé équilibré ***($\mu = 3{,}5$, $V = \frac{35}{12}$). Majorer la probabilité que la moyenne obtenue s'écarte de*** $3{,}5$ ***d'au moins*** $0{,}5$***.***

Solution. L'inégalité de concentration donne :

$$P\big(|M_{100} - 3{,}5| \geqslant 0{,}5\big) \leqslant \frac{35/12}{100 \times 0{,}5^2} = \frac{35}{300} = \frac{7}{60} \approx 0{,}117.$$

Remarque — Le facteur décisif est n

Multiplier la taille de l'échantillon par 100 divise la majoration par 100. C'est ce qui rend les sondages possibles.

4. Loi des grands nombres

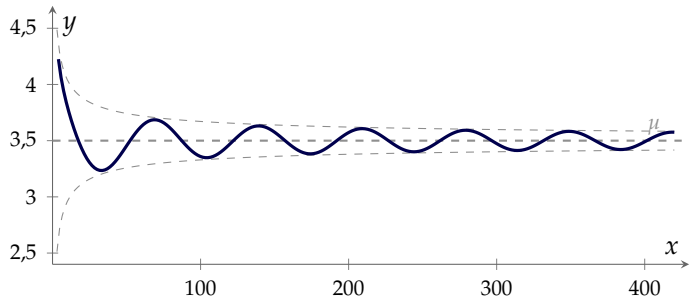


Figure 3 — La moyenne d'un échantillon se resserre autour de μ.

Théorème 14.5 — Loi des grands nombres

Avec les mêmes hypothèses, pour tout $\delta > 0$:

$$\lim_{n \to +\infty} P\big(|M_n - \mu| \geqslant \delta\big) = 0.$$

Interprétation. La moyenne d'un grand échantillon s'approche de l'espérance théorique avec une probabilité qui tend vers 1. C'est le fondement de l'estimation statistique.

Exercice d'application 5

Rédiger la démonstration de la loi des grands nombres pour le dé de l'application 4, avec $\delta = 0{,}5$**.**

Solution. Les lancers sont **indépendants et de même loi**, d'espérance 3,5 et de variance $\dfrac{35}{12}$. L'inégalité de concentration donne, pour tout n :

$$0 \leqslant P\big(|M_n - 3{,}5| \geqslant 0{,}5\big) \leqslant \frac{35/12}{n \times 0{,}25} = \frac{35}{3n}.$$

Comme $\dfrac{35}{3n} \to 0$, le théorème des gendarmes conclut : la probabilité tend vers 0.

Attention — Erreur à éviter

La loi des grands nombres ne dit **pas** que M_n finit par valoir μ, ni qu'un écart devient impossible. Elle dit que sa **probabilité** tend vers zéro. Un échantillon atypique reste concevable, il devient seulement rare.

Remarque — Rédaction attendue

Trois éléments : les variables sont indépendantes et de même loi, on applique l'inégalité de concentration, la majoration tend vers 0 car n est au dénominateur. Chacun vaut un point.

5. Finesse des majorations

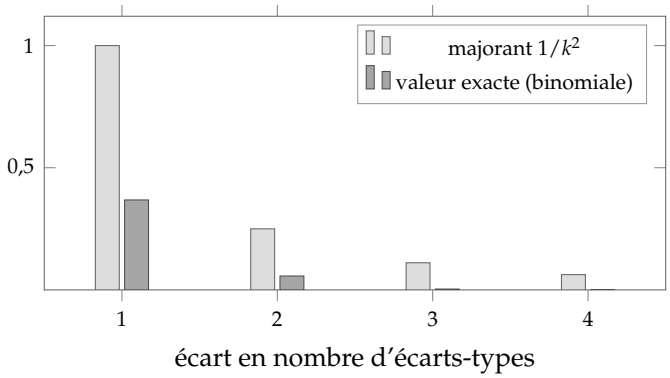


Figure 4 — Le majorant reste toujours au-dessus de la valeur exacte — ici pour $\mathcal{B}(100\,;0{,}5)$.

Ces inégalités sont **universelles** : elles valent pour toute loi vérifiant les hypothèses. Ce caractère général se paie par leur imprécision.

Exercice d'application 6

Pour X ***suivant*** $\mathcal{B}(100\,;0{,}5)$***, comparer la majoration de l'application 2 à la valeur exacte de*** $P\big(|X - 50| \geqslant 20\big)$***, qui est de l'ordre de*** $0{,}0001$**.**

Solution. Bienaymé-Tchebychev majore par 0,0625, contre une valeur exacte d'environ 0,0001 : la majoration est plus de 600 fois trop large.

Elle n'en reste pas moins **valable** : l'inégalité garantit sans estimer, pour toute loi de même espérance et de même variance — y compris la pire.

Remarque — Ce qu'il faut retenir

Une majoration n'est pas une valeur. Ces inégalités servent à **garantir** qu'une probabilité est petite, jamais à l'estimer. Quand l'énoncé demande « majorer », c'est le mot qui compte.

6. Dimensionner un échantillon

Exercice d'application 7

On veut estimer une proportion inconnue p par la fréquence M_n d'un sondage, en garantissant $P\big(|M_n - p| \geqslant 0{,}03\big) \leqslant 0{,}05$. Sachant que la variance d'une variable de Bernoulli vérifie $V = p(1-p) \leqslant \dfrac{1}{4}$, déterminer une taille n suffisante.

Solution. L'inégalité de concentration majore la probabilité par $\dfrac{V}{n\delta^2} \leqslant \dfrac{1/4}{n \times 0{,}03^2}$. Il suffit donc que

$$\frac{1/4}{n \times 0{,}0009} \leqslant 0{,}05 \iff n \geqslant \frac{0{,}25}{0{,}05 \times 0{,}0009} \approx 5\,555{,}6,$$

soit $n = 5\,556$ personnes interrogées.

Remarque — Bon réflexe

δ apparaît au carré : diviser la marge d'erreur par deux multiplie la taille requise par **quatre**. C'est la raison économique pour laquelle les sondages s'arrêtent à mille personnes.

Méthodes

Méthode 1 — Appliquer l'inégalité de Markov

1. Vérifier que X est à valeurs **positives ou nulles**.
2. Calculer ou relever $E(X)$.
3. Identifier le seuil a dans l'énoncé.
4. Écrire $P(X \geqslant a) \leqslant \dfrac{E(X)}{a}$ et conclure **dans le contexte**.

Exemple éclair. 200 visites en moyenne, seuil 1 000 : majoration 0,2 — c'est l'application 1.
Bon réflexe. La vérification du signe vaut un demi-point et prend une ligne. Un temps d'attente, un nombre d'objets, une durée sont positifs par nature : dites-le.

Méthode 2 — Appliquer Bienaymé-Tchebychev

1. Calculer $\mu = E(X)$ et $V = V(X)$.
2. Traduire l'énoncé en un écart $|X - \mu| \geqslant \delta$, en identifiant δ.
3. Appliquer $P\big(|X - \mu| \geqslant \delta\big) \leqslant \dfrac{V}{\delta^2}$.

4. Pour un événement de la forme « X hors de $]\mu-\delta\,;\mu+\delta[$ », c'est directement cette majoration.

Exemple éclair. $\mu = 50$, $V = 25$, $\delta = 20$: majoration $\dfrac{25}{400} = 0{,}0625$ — c'est l'application 2.

Erreur à éviter. Confondre variance et écart-type au dénominateur. C'est bien $\dfrac{V}{\delta^2}$, donc $\dfrac{\sigma^2}{\delta^2}$ — jamais $\dfrac{\sigma}{\delta^2}$.

Méthode 3 — Utiliser l'inégalité de concentration

1. Vérifier que les variables sont **indépendantes et de même loi**.
2. Identifier μ, V et la taille n de l'échantillon.
3. Traduire l'énoncé en un écart $|M_n - \mu| \geqslant \delta$.
4. Appliquer $P\big(|M_n - \mu| \geqslant \delta\big) \leqslant \dfrac{V}{n\delta^2}$.

Exemple éclair. 100 lancers du dé, $\delta = 0{,}5$: majoration $\dfrac{35}{300} \approx 0{,}117$ — c'est l'application 4.

Astuce. L'événement complémentaire « M_n dans l'intervalle » se minore alors par $1 - \dfrac{V}{n\delta^2}$. C'est souvent cette forme que l'énoncé attend.

Méthode 4 — Déterminer une taille d'échantillon

1. Écrire la condition demandée : $\dfrac{V}{n\delta^2} \leqslant \varepsilon$.
2. Isoler n : $n \geqslant \dfrac{V}{\varepsilon\,\delta^2}$.
3. Prendre le plus petit **entier** convenable.
4. Interpréter : « il faut interroger au moins n personnes ».

Exemple éclair. $\delta = 0{,}03$, $\varepsilon = 0{,}05$, $V \leqslant \dfrac{1}{4}$: $n = 5\,556$ — c'est l'application 7.

Bon réflexe. L'exposant 2 de δ est le cœur de la question : diviser la marge par deux multiplie la taille par quatre.

Méthode 5 — Rédiger une conclusion de loi des grands nombres

1. Rappeler que les variables sont indépendantes, de même loi, d'espérance μ et de variance V.
2. Écrire l'inégalité de concentration pour un δ fixé.
3. Observer que $\dfrac{V}{n\delta^2} \longrightarrow 0$ quand $n \to +\infty$.
4. Conclure : la probabilité que la moyenne s'écarte de μ de plus de δ tend vers 0.

Exemple éclair. Pour le dé : $0 \leqslant P \leqslant \dfrac{35}{3n} \to 0$, gendarmes — c'est l'application 5.

Erreur à éviter. Écrire « M_n tend vers μ ». C'est la **probabilité de l'écart** qui tend vers 0, pas la variable aléatoire elle-même. La formulation exacte est un point de barème.

Méthode 6 — Comparer une majoration à la valeur exacte

Quand ? L'énoncé demande de calculer une probabilité binomiale **et** de la majorer, puis de commenter.

1. Calculer la valeur exacte à la calculatrice.

2. Appliquer l'inégalité demandée.
3. Comparer les deux nombres, en donnant l'ordre de grandeur du rapport.
4. Conclure : l'inégalité est **universelle**, donc large ; elle garantit sans estimer.

Exemple éclair. $\mathcal{B}(100\,;0{,}5)$, $\delta = 10$: majorant 0,25, valeur exacte $\approx 0{,}057$, rapport 4,4.
Ce que le correcteur attend. Non pas « la majoration est mauvaise », mais « la majoration est valable pour **toute** loi vérifiant les hypothèses, ce qui explique qu'elle soit bien au-dessus de la valeur exacte pour une loi particulière ».

Erreurs fréquentes et bons réflexes

Erreur	Bon réflexe
Appliquer Markov à une variable pouvant être négative	L'hypothèse $X \geqslant 0$ est indispensable
Écrire $\dfrac{V(X)}{\delta}$ au lieu de $\dfrac{V(X)}{\delta^2}$	Le dénominateur est au **carré**
Utiliser σ au lieu de σ^2 au numérateur	C'est la **variance**
Annoncer un majorant supérieur à 1 comme une information	Toute probabilité est déjà $\leqslant 1$
Croire que Bienaymé-Tchebychev donne la valeur exacte	C'est une **majoration**, souvent large
Dire que M_n tend vers μ	C'est la **probabilité de l'écart** qui tend vers 0
Oublier $\dfrac{1}{n}$ dans l'inégalité de concentration	C'est ce facteur qui fait toute la loi des grands nombres

Module Python — majoration contre valeur exacte

```
# Majoration de Bienayme-Tchebychev contre valeur exacte
def combinaison(n, k):
  resultat = 1
  for i in range(k):
     resultat = resultat * (n - i) // (i + 1)
  return resultat
def proba_binomiale(n, p, k):
  return combinaison(n, k) * p ** k * (1 - p) ** (n - k)
def proba_exacte_ecart(n, p, delta):
  # P(|X - np| >= delta)
  mu = n * p
  total = 0
  for k in range(n + 1):
     if abs(k - mu) >= delta:
        total = total + proba_binomiale(n, p, k)
  return total
def majorant(n, p, delta):
  return n * p * (1 - p) / (delta * delta)
liste_n = [100, 100, 400, 200]
liste_p = [0.5, 0.5, 0.5, 0.4]
liste_delta = [10, 15, 30, 20]
print("B(n ; p) :  majorant de Bienayme-Tchebychev et valeur exacte")
for i in range(len(liste_n)):
  n = liste_n[i]
  p = liste_p[i]
  delta = liste_delta[i]
  m = majorant(n, p, delta)
  e = proba_exacte_ecart(n, p, delta)
  print("n =", n, " p =", p, " delta =", delta)
  print(" majorant :", round(m, 5), " exact :", round(e, 6))
  print(" rapport :", round(m / e, 1))
```

Trace d'exécution.

```
B(n ; p) :  majorant de Bienayme-Tchebychev et valeur exacte
n = 100 p = 0.5 delta = 10
majorant :  0.25 exact :  0.056888
rapport :  4.4
n = 100 p = 0.5 delta = 15
majorant :  0.11111 exact :  0.003518
rapport :  31.6
n = 400 p = 0.5 delta = 30
majorant :  0.11111 exact :  0.003129
rapport :  35.5
n = 200 p = 0.4 delta = 20
majorant :  0.12 exact :  0.004766
rapport :  25.2
```

Interprétation mathématique. Le majorant est **toujours supérieur** à la valeur exacte : l'inégalité est vérifiée dans les quatre cas. ✓
Mais le rapport se dégrade fortement quand δ s'éloigne de l'écart-type : il vaut 4,4 pour un écart de deux écarts-types ($n = 100$, $\delta = 10$, $\sigma = 5$) et monte à 35,5 pour un écart de trois écarts-types ($n = 400$, $\delta = 30$, $\sigma = 10$). C'est le prix de l'universalité — l'inégalité doit rester vraie pour la pire loi possible, et la binomiale est très loin d'être ce pire cas.

Attention

Ce que cela signifie en pratique. Une taille d'échantillon calculée par Bienaymé-Tchebychev est **surdimensionnée**, souvent d'un facteur 10 ou plus. Elle a l'avantage décisif de ne demander aucune hypothèse : on l'utilise quand on ne connaît rien de la loi, ou quand on veut une garantie absolue.