# Plaquette 4 – Inégalités de concentration et loi faible des grands nombres

## Espérance et variance — Prépa MPSI / MP2I (1re année)

Corrections détaillées

**Boîte à outils**

**Concentration.**

— **Markov** : si $X \geq 0$ et $a > 0$, $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}(X)}{a}$.

— **Bienaymé-Tchebychev** : pour $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{V}(X)}{\varepsilon^2}$$

— **Moyenne empirique** de $X_1, \ldots, X_n$ indépendantes de même loi (espérance $m$, variance $\sigma^2$) : $\overline{X}_n = \frac{1}{n}\sum X_i$ vérifie $\mathbb{E}(\overline{X}_n) = m$ et $\mathrm{V}(\overline{X}_n) = \frac{\sigma^2}{n}$.

— **Loi faible des grands nombres** :

$$\mathbb{P}(|\overline{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \underset{n \to +\infty}{\longrightarrow} 0$$

— Pour une **fréquence** ($X_i$ de Bernoulli) : $\sigma^2 = p(1-p) \leq \frac{1}{4}$.

— **Markov exponentiel** : $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}(\mathrm{e}^{tX})}{\mathrm{e}^{ta}}$ pour tout $t > 0$.

## Correction de l’exercice 1 – L’inégalité de Markov en pratique

a) Comme $X \geq 0$, on a $X \geq a\mathbf{1}_{\{X \geq a\}}$ (si $X \geq a$, c’est $X \geq a$ ; sinon, $X \geq 0$). Par croissance de l’espérance :

$$\mathbb{E}(X) \geq a\,\mathbb{P}(X \geq a)$$

b) En tirant un salarié au hasard : $\mathbb{P}(X \geq 10\,000) \leq \dfrac{2\,500}{10\,000} = 25\,\%$.

c) Égalité si $X \in \{0, a\}$ presque sûrement : 25 % des salariés à 10 000 € et 75 % à 0 € donnent bien une moyenne de 2 500 €. La borne est donc optimale sans autre information.

d) $\dfrac{2\,500}{2\,000} > 1$ : la borne est sans intérêt. Markov n’est utile que pour des seuils nettement supérieurs à la moyenne.

**Réponse.** $\mathbb{P}(X \geq 10\,000) \leq 25\,\%$, borne atteinte (25 % à 10 000, 75 % à 0) ; inutile pour $a <$ moyenne.

## Correction de l’exercice 2 – Tchebychev contre calcul exact

a) $|X - 5| \geq 3 \Leftrightarrow X \leq 2$ ou $X \geq 8$ ; par symétrie :

$$\mathbb{P} = 2 \times \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2}}{1\,024} = \frac{2 \times 56}{1\,024} \approx 0{,}109$$

b) $\mathrm{V}(X) = 10 \times \frac{1}{4} = 2{,}5$ : $\mathbb{P}(|X - 5| \geq 3) \leq \dfrac{2{,}5}{9} \approx 0{,}278$.

c) Exact : $2 \times \dfrac{1 + 10 + 45 + 120}{1\,024} = \dfrac{352}{1\,024} \approx 0{,}344$ ; borne : $\dfrac{2{,}5}{4} = 0{,}625$. La borne est valable mais 2 à 3 fois trop grande : Tchebychev ne connaît que la variance et doit être vrai pour

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

toutes les lois de même variance.

**Réponse.** Exact ≈ 0,109 contre ≤ 0,278 ; exact ≈ 0,344 contre ≤ 0,625.

## Correction de l'exercice 3 – Somme de cent dés

a) Linéarité et indépendance : $\mathbb{E}(S) = 100 \times 3{,}5 = 350$, $\mathrm{V}(S) = 100 \times \frac{35}{12} = \frac{875}{3} \approx 291{,}7$.

b) Bienaymé-Tchebychev :

$$\mathbb{P}(|S - 350| \geq 50) \leq \frac{291{,}7}{2\,500} \approx 0{,}117$$

c) $\frac{291{,}7}{\varepsilon^2} \leq 0{,}05 \Leftrightarrow \varepsilon^2 \geq 5\,833 \Leftrightarrow \varepsilon \geq 76{,}4$. Avec probabilité au moins 95 %, $S$ est entre 274 et 426 (la vraie fourchette, plus étroite, est d'environ ±34).

**Réponse.** $\mathbb{E}(S) = 350$, $\mathrm{V}(S) \approx 291{,}7$ ; ≤ 0,117 ; $\varepsilon \approx 76{,}4$.

## Correction de l'exercice 4 – La loi faible des grands nombres pour une fréquence

a) $F_n = \frac{S}{n}$ avec $S \sim B(n, p)$ :

$$\mathbb{E}(F_n) = p, \qquad \mathrm{V}(F_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

b) Tchebychev, puis $p(1-p) \leq \frac{1}{4}$ : $\mathbb{P}(|F_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$.

c) Pour $\varepsilon$ fixé, la borne tend vers 0 quand $n \to +\infty$ : la fréquence observée se rapproche de la probabilité avec une probabilité qui tend vers 1. C'est la justification mathématique de l'interprétation « fréquentiste » des probabilités.

**Réponse.** $\mathbb{E}(F_n) = p$, $\mathrm{V}(F_n) = \frac{p(1-p)}{n}$ ; borne $\frac{1}{4n\varepsilon^2} \to 0$.

## Correction de l'exercice 5 – Taille d'un échantillon de sondage

a) On impose $\frac{1}{4n\varepsilon^2} \leq \delta$ avec $\varepsilon = 0{,}05$, $\delta = 0{,}05$ :

$$n \geq \frac{1}{4\varepsilon^2\delta} = \frac{1}{4 \times 0{,}0025 \times 0{,}05} = 2\,000$$

b) $\varepsilon = 0{,}01$ : $n \geq \frac{1}{4 \times 0{,}0001 \times 0{,}05} = 50\,000$.

c) $n$ est proportionnel à $\frac{1}{\varepsilon^2}$ : diviser la marge d'erreur par 5 multiplie la taille de l'échantillon par 25. (En pratique, la loi normale permet des tailles plus petites, environ 385 au lieu de 2 000 pour a).)

**Réponse.** $n \geq 2\,000$ ; $n \geq 50\,000$ ; $n \propto \frac{1}{\varepsilon^2}$.

## Correction de l'exercice 6 – Moyenne de lancers de dés

a) $\mathbb{E}(\overline{X}_n) = 3{,}5$, $\mathrm{V}(\overline{X}_n) = \frac{35}{12n}$.

b) $\frac{35}{12n \times 0{,}01} \leq 0{,}05 \Leftrightarrow n \geq \frac{35}{12 \times 0{,}0005} \approx 5\,833{,}3$ : $n = 5\,834$.

c) Pour un dé équilibré, $\mathbb{P}(|\overline{X}_{1000} - 3{,}5| \geq 0{,}3) \leq \frac{35}{12 \times 1\,000 \times 0{,}09} \approx 0{,}032$. Un écart de 0,3 a moins de 3,2 % de chances de se produire : le dé est suspect (au seuil de 5 %).

**Réponse.** $\mathrm{V}(\overline{X}_n) = \frac{35}{12n}$ ; $n = 5\,834$ ; écart 0,3 : probabilité ≤ 3,2 %, dé suspect.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

## Correction de l'exercice 7 – Majorer grâce au second moment

a) $\{|X| \geq a\} = \{X^2 \geq a^2\}$ et $X^2 \geq 0$ ; Markov :

$$\mathbb{P}(|X| \geq a) = \mathbb{P}(X^2 \geq a^2) \leq \frac{\mathbb{E}(X^2)}{a^2}$$

b) Avec $Y = X - \mathbb{E}(X)$ : $\mathbb{E}(Y^2) = \mathrm{V}(X)$, d'où $\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathrm{V}(X)}{\varepsilon^2}$.

c) $\mathbb{P}(|X| \geq 1) \leq \frac{0{,}04}{1} = 4\,\%$ (erreur de 5 écarts types).

**Réponse.** $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(X^2)}{a^2}$ ; c) $\leq 4\,\%$.

## Correction de l'exercice 8 – Contrôle qualité à grande échelle

a) $X \sim B(10\,000;\, 0{,}01)$ : $\mathbb{E}(X) = 100$, $\mathrm{V}(X) = 99$, $\sigma \approx 9{,}95$.

b) Markov : $\mathbb{P}(X \geq 150) \leq \frac{100}{150} \approx 0{,}67$ (inutile). Tchebychev :
$\{X \geq 150\} \subset \{|X - 100| \geq 50\}$, donc

$$\mathbb{P}(X \geq 150) \leq \frac{99}{2\,500} \approx 0{,}040$$

c) On veut $\mathbb{P}(X > 100 + \varepsilon) \leq 0{,}01$ ; il suffit que $\frac{99}{\varepsilon^2} \leq 0{,}01$, soit $\varepsilon \geq \sqrt{9\,900} \approx 99{,}5$. Un stock de 200 pièces suffit (borne pessimiste : en réalité, 125 environ suffiraient).

**Réponse.** $100 \pm 9{,}95$ ; Markov $\leq 0{,}67$, Tchebychev $\leq 0{,}04$ ; stock de 200 pièces.

## Correction de l'exercice 9 – Une borne exponentielle (inégalité de Chernoff)

a) $\mathbb{E}(t^X) = G_X(t) = \left(\frac{1}{2} + \frac{t}{2}\right)^n$ ; en $t = 3$ : $2^n$.

b) $\{X \geq \frac{3n}{4}\} = \{3^X \geq 3^{3n/4}\}$ (croissance de $t \mapsto 3^t$) et $3^X > 0$ :

$$\mathbb{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{\mathbb{E}(3^X)}{3^{3n/4}} = \frac{2^n}{3^{3n/4}} = \left(\frac{2}{3^{3/4}}\right)^n$$

c) $n = 100$ : borne exponentielle $\approx 2 \times 10^{-6}$. Tchebychev : $\mathbb{P}(|X - 50| \geq 25) \leq \frac{25}{625} = 0{,}04$.
La borne exponentielle est **vingt mille fois** meilleure : elle exploite toute la loi (via $\mathbb{E}(3^X)$) et décroît exponentiellement en $n$.

**Réponse.** $\mathbb{E}(3^X) = 2^n$ ; $\mathbb{P}(X \geq \frac{3n}{4}) \leq (\frac{2}{3^{3/4}})^n \approx 2 \times 10^{-6}$ pour $n = 100$, contre 0,04.

## Correction de l'exercice 10 – Somme de variables de Bernoulli de paramètres différents

a) $\mathbb{E}(S) = \sum p_i = n\overline{p}$ ; par indépendance, $\mathrm{V}(S) = \sum p_i(1 - p_i) \leq \frac{n}{4}$.

b) $\mathrm{V}\left(\frac{S}{n}\right) = \frac{\mathrm{V}(S)}{n^2} \leq \frac{1}{4n}$ et Tchebychev :

$$\mathbb{P}\left(\left|\frac{S}{n} - \overline{p}\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

c) $n = 100$, $\varepsilon = 0{,}15$ : $\mathbb{P}(|S - 70| \geq 15) \leq \frac{1}{4 \times 100 \times 0{,}0225} \approx 0{,}111$. La loi faible vaut encore sans loi identique.

**Réponse.** $\mathrm{V}(S) = \sum p_i(1 - p_i) \leq \frac{n}{4}$ ; c) $\leq 0{,}111$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

## Correction de l'exercice 11 – Quand Tchebychev ne dit rien

a) $\mathrm{V}(X) = \frac{n}{4}$ ; la borne $\frac{n}{4\varepsilon^2} \geq 1 \Leftrightarrow \varepsilon \leq \frac{\sqrt{n}}{2} = \sigma(X)$.

b) Tchebychev ne dit rien pour des écarts inférieurs à un écart type : les fluctuations naturelles de $X$ sont de l'ordre de $\sigma = \frac{\sqrt{n}}{2}$. Les écarts de plusieurs écarts types sont rares.

c) $\sigma = 10$. $\varepsilon = 20$ $(2\sigma)$ : borne $\frac{100}{400} = 0{,}25$. $\varepsilon = 40$ $(4\sigma)$ : $\frac{100}{1\,600} \approx 0{,}063$.

$$\mathbb{P}(|X - \mathbb{E}X| \geq k\sigma) \leq \frac{1}{k^2}$$

**Réponse.** Borne inutile pour $\varepsilon \leq \sigma$ ; $n = 400$ : $\sigma = 10$, bornes 0,25 et 0,063.

## Correction de l'exercice 12 – Estimer une probabilité par simulation

a) Script :

```
from random import randint

def frequence(n):
    succes = 0
    for _ in range(n):
        s = randint(1, 6) + randint(1, 6) + randint(1, 6)
        if s > 12:
            succes = succes + 1
    return succes / n
```

b) $\frac{1}{4n\varepsilon^2} \leq \delta$ avec $\varepsilon = 0{,}01$, $\delta = 0{,}01$ :

$$n \geq \frac{1}{4 \times 10^{-4} \times 10^{-2}} = 250\,000$$

c) $p = \frac{56}{216} \approx 0{,}259$ (56 triplets de somme 13 à 18 : $21 + 15 + 10 + 6 + 3 + 1$). La simulation est une méthode générale (Monte-Carlo), mais la précision coûte cher : elle ne progresse qu'en $\frac{1}{\sqrt{n}}$.

**Réponse.** Programme Monte-Carlo ; $n \geq 250\,000$ ; $p = \frac{56}{216} \approx 0{,}259$.

## Correction de l'exercice 13 – Moyenne empirique : espérance et variance

a) Linéarité : $\mathbb{E}(\overline{X}_n) = \frac{1}{n} \times nm = m$. Variance d'une somme de variables indépendantes :

$$\mathrm{V}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{V}(X_i) = \frac{\sigma^2}{n}$$

b) Pour annuler les covariances dans la variance de la somme (l'indépendance deux à deux suffit). Si toutes les variables sont égales, $\overline{X}_n = X_1$ et $\mathrm{V}(\overline{X}_n) = \sigma^2$ : moyenner des copies identiques n'apporte rien.

c) Tchebychev : $\mathbb{P}(|\overline{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$.

**Réponse.** $\mathbb{E}(\overline{X}_n) = m$, $\mathrm{V}(\overline{X}_n) = \frac{\sigma^2}{n}$ (indépendance) ; LFGN par Tchebychev.

## Correction de l'exercice 14 – Markov et les valeurs extrêmes

a) $Y = 10 - X \geq 0$ et $\mathbb{E}(Y) = 2$. $X \leq 5 \Leftrightarrow Y \geq 5$, donc

$$\mathbb{P}(X \leq 5) = \mathbb{P}(Y \geq 5) \leq \frac{2}{5}$$

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

b) $\mathbb{P}(X \geq 6) = 1 - \mathbb{P}(X \leq 5) \geq 1 - \frac{2}{5} = \frac{3}{5} \geq \frac{1}{2}$.

c) Égalité dans Markov si $Y \in \{0, 5\}$ : $\mathbb{P}(Y = 5) = \frac{2}{5}$, $\mathbb{P}(Y = 0) = \frac{3}{5}$, soit $\mathbb{P}(X = 5) = \frac{2}{5}$ et $\mathbb{P}(X = 10) = \frac{3}{5}$, d'espérance $2 + 6 = 8$.

**Réponse.** $\mathbb{P}(X \leq 5) \leq \frac{2}{5}$ ; $\mathbb{P}(X \geq 6) \geq \frac{3}{5}$ ; égalité pour $X \in \{5, 10\}$ avec $\frac{2}{5}$, $\frac{3}{5}$.

## Correction de l'exercice 15 – Synthèse : la loi des grands nombres et l'assurance

a) $X \sim B(n;\, 0{,}01)$ ; $\mathbb{E}(R) = 120n - 10\,000 \times 0{,}01n = 20n$ : gain moyen de 20 € par client.

b) $R < 0 \Leftrightarrow X > 0{,}012n \Leftrightarrow X - 0{,}01n > 0{,}002n$, inclus dans $\{|X - \mathbb{E}X| \geq 0{,}002n\}$. Avec $\mathrm{V}(X) = 0{,}0099n$ :

$$\mathbb{P}(R < 0) \leq \frac{0{,}0099n}{(0{,}002n)^2} = \frac{0{,}0099}{4 \times 10^{-6} n} = \frac{2\,475}{n}$$

c) $\frac{2\,475}{n} \leq 0{,}01 \Leftrightarrow n \geq 247\,500$ clients.

d) Avec peu de clients, un seul sinistre suffit à ruiner l'exercice ; la loi des grands nombres garantit qu'avec beaucoup de clients, la fréquence des sinistres est proche de 1 %. La marge (20 € au-dessus du coût moyen) est indispensable : avec une prime de 100 €, $\mathbb{E}(R) = 0$ et la compagnie serait en perte une année sur deux environ.

**Réponse.** $\mathbb{E}(R) = 20n$ ; $\mathbb{P}(R < 0) \leq \frac{2475}{n}$ ; $n \geq 247\,500$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).