

Corrigé : Sommes de variables aléatoires et concentration Fiche Défi**D1.**

1. Par définition, $V = \sum_{i=1}^r p_i(x_i - \mu)^2$, somme de termes positifs ou nuls. En ne gardant que les indices de I , on retire des termes positifs ou nuls : $V \geq \sum_{i \in I} p_i(x_i - \mu)^2$. Pour $i \in I$, $|x_i - \mu| \geq \delta$ donc $(x_i - \mu)^2 \geq \delta^2$ et $p_i(x_i - \mu)^2 \geq p_i\delta^2$. Ainsi :

$$V \geq \sum_{i \in I} p_i(x_i - \mu)^2 \geq \delta^2 \sum_{i \in I} p_i = \delta^2 P(|X - \mu| \geq \delta),$$

car l'événement $(|X - \mu| \geq \delta)$ est la réunion des événements disjoints $(X = x_i)$, $i \in I$. En divisant par $\delta^2 > 0$: $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$.

2. Comme $\sigma > 0$, $\delta = k\sigma > 0$ est admissible : $P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$. Pour $k = 2$: au plus $\frac{1}{4}$; pour $k = 3$: au plus $\frac{1}{9} \approx 0,111$.

3. Si $V = 0$, l'inégalité donne $0 \leq P(|X - \mu| \geq \delta) \leq \frac{0}{\delta^2} = 0$ pour tout $\delta > 0$, donc cette probabilité est nulle.

Si tous les x_i valent μ , alors $X = \mu$ et $P(X = \mu) = 1$. Sinon, soit δ_0 le plus petit des écarts non nuls $|x_i - \mu|$ (il y en a un nombre fini). Tout résultat avec $X \neq \mu$ vérifie $|X - \mu| \geq \delta_0$, donc $P(X \neq \mu) \leq P(|X - \mu| \geq \delta_0) = 0$. Ainsi $P(X = \mu) = 1$.

Conclusion : $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$; $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$; **une variable de variance nulle est presque sûrement constante.**

D2.

1. Par indépendance, $P(X = x_i \text{ et } Y = y_j) = P(X = x_i)P(Y = y_j)$. Donc

$$E(XY) = \sum_{i,j} x_i y_j P(X = x_i)P(Y = y_j) = \left(\sum_{i=1}^r x_i P(X = x_i) \right) \left(\sum_{j=1}^s y_j P(Y = y_j) \right) = E(X)E(Y).$$

2. On a $(X + Y)^2 = X^2 + 2XY + Y^2$. Par linéarité :

$$V(X + Y) = E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2.$$

En regroupant : $V(X + Y) = V(X) + V(Y) + 2[E(XY) - E(X)E(Y)]$. D'après la question 1, le crochet est nul : $V(X + Y) = V(X) + V(Y)$.

3. Comme X et $-Y$ sont indépendantes, la question 2 donne $V(X - Y) = V(X) + V(-Y) = V(X) + (-1)^2 V(Y) = V(X) + V(Y)$.

4. $S = X + Y$ prend les valeurs 0, 1, 2 avec $P(S = 0) = (1-p)^2$, $P(S = 1) = 2p(1-p)$, $P(S = 2) = p^2$. Alors $E(S) = 2p(1-p) + 2p^2 = 2p$ et $E(S^2) = 2p(1-p) + 4p^2$, donc $V(S) = 2p(1-p) + 4p^2 - 4p^2 = 2p(1-p) = V(X) + V(Y)$.

5. Non. $E(X) = \frac{-2 - 1 + 1 + 2}{4} = 0$, donc $E(X)E(Y) = 0$. $XY = X^3$ vaut -8, -1, 1, 8 avec la probabilité $\frac{1}{4}$ chacune : $E(XY) = \frac{-8 - 1 + 1 + 8}{4} = 0$. Les deux nombres sont égaux. Pourtant $P(X = 2 \text{ et } Y = 4) = P(X = 2) = \frac{1}{4}$ alors que $P(X = 2)P(Y = 4) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$: X et Y ne sont pas indépendantes.

Conclusion : l'indépendance entraîne $E(XY) = E(X)E(Y)$ et $V(X \pm Y) = V(X) + V(Y)$, mais $E(XY) = E(X)E(Y)$ n'entraîne pas l'indépendance.

D3.

1. Les répartitions où la personne i retrouve son chapeau correspondent aux $(n-1)!$ façons de répartir les $n-1$ autres chapeaux : $P(X_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$. Comme $N = X_1 + \dots + X_n$, la linéarité (sans indépendance) donne $E(N) = n \times \frac{1}{n} = 1$. En moyenne, une personne exactement retrouve son chapeau, quel que soit n .
2. $X_i X_j = 1$ si et seulement si i et j retrouvent leur chapeau : $(n-2)!$ répartitions, donc $E(X_i X_j) = P(X_i = 1 \text{ et } X_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$. Or $E(X_i)E(X_j) = \frac{1}{n^2} \neq \frac{1}{n(n-1)}$: les variables X_i ne se comportent pas comme des variables indépendantes, et l'additivité de la variance ne peut pas être utilisée pour N (d'où le calcul de $E(N^2)$ à la question suivante).
3. $X_i^2 = X_i$ (valeurs 0 et 1), donc $E(X_i^2) = \frac{1}{n}$. Il y a $n(n-1)$ couples ordonnés (i, j) avec $i \neq j$:

$$E(N^2) = n \times \frac{1}{n} + n(n-1) \times \frac{1}{n(n-1)} = 1 + 1 = 2.$$

Alors $V(N) = E(N^2) - E(N)^2 = 2 - 1 = 1$ pour tout $n \geq 2$.

4. Pour $n = 4$: si trois personnes retrouvent leur chapeau, le quatrième chapeau est celui de la dernière personne, donc $N = 4$: $P(N = 3) = 0$. Un seul cas où chacune retrouve le sien : $P(N = 4) = \frac{1}{24}$. Pour $N = 2$: on choisit les deux personnes qui retrouvent leur chapeau ($\binom{4}{2} = 6$ façons), et les deux autres doivent échanger leurs chapeaux (une seule façon) : $P(N = 2) = \frac{6}{24} = \frac{1}{4}$. Pour $N = 1$: 4 choix de la personne, puis les 3 autres chapeaux sont répartis sans point fixe (deux façons : les deux permutations circulaires) : $P(N = 1) = \frac{8}{24} = \frac{1}{3}$. Enfin $P(N = 0) = 1 - \frac{1}{24} - \frac{1}{4} - \frac{1}{3} = \frac{9}{24} = \frac{3}{8}$.

k	0	1	2	3	4
$P(N = k)$	$\frac{3}{8}$	$\frac{1}{3}$	$\frac{1}{4}$	0	$\frac{1}{24}$

$$E(N) = \frac{1}{3} + \frac{2}{4} + \frac{4}{24} = \frac{1}{3} + \frac{1}{2} + \frac{1}{6} = 1; \quad E(N^2) = \frac{1}{3} + 4 \times \frac{1}{4} + 16 \times \frac{1}{24} = \frac{1}{3} + 1 + \frac{2}{3} = 2, \text{ donc } V(N) = 1 : \text{cohérent avec les questions 1 et 3.}$$

5. Avec $\mu = 1$ et $V = 1$: $N \geq 6 \Rightarrow |N - 1| \geq 5$, donc $P(N \geq 6) \leq \frac{1}{25} = 0,04$.

Conclusion : $E(N) = 1$ et $V(N) = 1$ pour tout $n \geq 2$ (variables dépendantes) ; pour $n = 10$, $P(N \geq 6) \leq 0,04$.

D4.

1. Par linéarité : $E(Z_a) = aE(X) + (1-a)E(Y) = am + (1-a)m = m$: le résultat est toujours sans biais. Par indépendance : $V(Z_a) = a^2V(X) + (1-a)^2V(Y) = 4a^2 + (1-a)^2 = 5a^2 - 2a + 1$.
2. Soit $f(a) = 5a^2 - 2a + 1$ sur $[0; 1]$. $f'(a) = 10a - 2$ s'annule en $a = \frac{1}{5}$; $f' < 0$ avant, $f' > 0$ après : f admet un minimum en $\frac{1}{5} \in [0; 1]$, et $f\left(\frac{1}{5}\right) = \frac{1}{5} - \frac{2}{5} + 1 = \frac{4}{5}$.
3. $V(Z_{1/5}) = 0,8 < 1 = V(Y) < 4 = V(X)$. Combiner les deux mesures, avec un poids plus fort sur la plus précise, donne un résultat plus précis que le meilleur des deux instruments pris seul.
4. Pour $a = \frac{1}{5}$: $P(|Z - m| \geq 2) \leq \frac{0,8}{4} = 0,2$, donc $P(|Z - m| < 2) \geq 0,8$. Avec X seul : $P(|X - m| \geq 2) \leq \frac{4}{4} = 1$, donc la minoration obtenue est 0 : elle est sans intérêt.
5. $V(Z_a) = \sigma_1^2 a^2 + \sigma_2^2 (1-a)^2$, de dérivée $2\sigma_1^2 a - 2\sigma_2^2 (1-a) = 2(\sigma_1^2 + \sigma_2^2)a - 2\sigma_2^2$, qui s'annule en $a^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \in]0; 1[$ (la dérivée est croissante : c'est un minimum). Avec $1 - a^* = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$:

$$V(Z_{a^*}) = \frac{\sigma_1^2 \sigma_2^4 + \sigma_2^2 \sigma_1^4}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2 \sigma_2^2 (\sigma_2^2 + \sigma_1^2)}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Pour $\sigma_1^2 = 4$ et $\sigma_2^2 = 1$, on retrouve $a^* = \frac{1}{5}$ et $\frac{4}{5}$.

Conclusion : $a = \frac{1}{5}$ minimise la variance, qui vaut alors 0,8, inférieure aux deux variances initiales.

D5.

1. $E(F_n) = 0,5$ et $V(F_n) = \frac{0,25}{n}$. Avec $\delta = 0,05$: $P(A_n) \leq b_n = \frac{0,25}{n \times 0,0025} = \frac{100}{n}$.
2. $F_n = \frac{X_n}{n}$ et $|F_n - 0,5| \geq 0,05 \iff \left| \frac{2X_n - n}{2n} \right| \geq \frac{1}{20} \iff 20|2X_n - n| \geq 2n \iff 10|2X_n - n| \geq n$.
3. Les 2^n suites de lancers sont équiprobables et $\binom{n}{k}$ d'entre elles ont exactement k « pile ». La somme dénombre les suites pour lesquelles $10|2k - n| \geq n$ (question 2), donc leur nombre divisé par 2^n est $P(A_n)$.

n	100	400	1 000	2 000
$b_n = \frac{100}{n}$	1	0,25	0,1	0,05
$P(A_n)$	0,368	0,0510	0,00173	$8,46 \times 10^{-6}$
$b_n/P(A_n)$	$\approx 2,7$	$\approx 4,9$	≈ 58	$\approx 5 900$

4. La borne b_n décroît seulement comme $\frac{1}{n}$: elle est divisée par 2 quand n double. Le risque exact décroît bien plus vite, comme le montre le rapport $b_n/P(A_n)$ qui explose (de 2,7 à environ 5 900). Pour $n = 100$, $b_n = 1$ n'apporte aucune information alors que le risque est déjà de 37 %. On utilise tout de même l'inégalité car elle ne demande que l'espérance et la variance, vaut pour toute loi et donne une garantie rigoureuse et simple, alors que la loi exacte n'est pas toujours calculable à la main.

Conclusion : $P(A_n) \leq \frac{100}{n}$: cette borne est valable, mais très pessimiste pour la loi binomiale (par exemple 0,05 contre $8,46 \times 10^{-6}$ pour $n = 2 000$).

D6. Partie A

1. Pour le livreur A, on répète 50 fois de façon identique et indépendante une épreuve de Bernoulli de probabilité de succès « colis en retard » 0,08 : $N \sim \mathcal{B}(50; 0,08)$. De même $M \sim \mathcal{B}(80; 0,05)$.
 $E(N) = 4$, $V(N) = 50 \times 0,08 \times 0,92 = 3,68$; $E(M) = 4$, $V(M) = 80 \times 0,05 \times 0,95 = 3,8$.
2. Par linéarité : $E(T) = 4 + 4 = 8$. Comme N et M sont indépendantes : $V(T) = 3,68 + 3,8 = 7,48$ et $\sigma(T) = \sqrt{7,48} \approx 2,73$.
3. Si $T \sim \mathcal{B}(n, q)$, alors $nq = 8$ et $nq(1 - q) = 7,48$. En divisant : $1 - q = \frac{7,48}{8} = 0,935$, donc $q = 0,065$ et $n = \frac{8}{0,065} = \frac{1 600}{13} \approx 123,08$, qui n'est pas un entier. C'est impossible : T ne suit aucune loi binomiale.

Partie B

4. Avec $\mu = 8$: $T \geq 20 \Rightarrow T - 8 \geq 12 \Rightarrow |T - 8| \geq 12$. D'après l'inégalité de Bienaymé-Tchebychev avec $\delta = 12$: $P(T \geq 20) \leq \frac{7,48}{144} \approx 0,0519 \leq 0,052$.
5. T est entier : $3 \leq T \leq 13 \iff |T - 8| \leq 5 \iff |T - 8| < 6$. Avec $\delta = 6$: $P(|T - 8| \geq 6) \leq \frac{7,48}{36} \approx 0,2078$. Par événement contraire : $P(3 \leq T \leq 13) \geq 1 - 0,2078 = 0,7922$, soit au moins 0,79.

Partie C

6. Soit $f(p) = p(1 - p) = p - p^2$. $f'(p) = 1 - 2p > 0$ sur $[0; 0,1]$: f est croissante, donc $f(p) \leq f(0,1) = 0,1 \times 0,9 = 0,09$.
7. Le nombre de retards suit $\mathcal{B}(n, p)$: $E(F_n) = p$ et $V(F_n) = \frac{p(1 - p)}{n} \leq \frac{0,09}{n}$. Avec $\delta = 0,01$:
 $P(|F_n - p| \geq 0,01) \leq \frac{0,09}{n \times 0,0001} = \frac{900}{n}$. On veut $\frac{900}{n} \leq 0,05 \iff n \geq 18 000$. Le plus petit entier est $n = 18 000$. Sans l'information $p \leq 0,1$, on majore par $\frac{1}{4}$: $\frac{2 500}{n} \leq 0,05 \iff n \geq 50 000$.

Partie D

8. import random

```
def retards(n, p):
    return sum(1 for _ in range(n) if random.random() < p)

def estime(N):
    c = 0
    for _ in range(N):
        if retards(50, 0.08) + retards(80, 0.05) >= 20:
            c = c + 1
    return c / N
```

D'après la question 4, la fréquence simulée doit être inférieure à 0,052 avec une probabilité élevée. En fait, un calcul exact de la loi de T donne $P(T \geq 20) \approx 1,4 \times 10^{-4}$: en moyenne, `estime(10000)` compte environ 1 ou 2 journées sur 10 000 et renvoie donc une valeur de l'ordre de 10^{-4} (souvent 0, parfois 0,0001 ou 0,0002), bien en dessous de la borne 0,052.

Conclusion : $E(T) = 8$, $V(T) = 7,48$, T **non binomiale**; $P(T \geq 20) \leq 0,052$; $P(3 \leq T \leq 13) \geq 0,79$; $n = 18\,000$ **colis suffisent (contre 50 000 sans information sur p)**.

D7.

1. $E(X) = -3 \times 0,08 + 0 + 3 \times 0,08 = 0$ et $V(X) = 9 \times 0,08 + 9 \times 0,08 = 1,44$. La majoration pour $\delta = 3$ vaut $\frac{1,44}{9} = 0,16$, et $P(|X| \geq 3) = 0,08 + 0,08 = 0,16$: il y a égalité.
2. $E(X) = \frac{q}{2}(\mu - \delta) + (1 - q)\mu + \frac{q}{2}(\mu + \delta) = \mu$. $V(X) = \frac{q}{2}\delta^2 + 0 + \frac{q}{2}\delta^2 = q\delta^2$. Enfin $|X - \mu| \geq \delta$ si et seulement si $X \neq \mu$: $P(|X - \mu| \geq \delta) = \frac{q}{2} + \frac{q}{2} = q = \frac{V(X)}{\delta^2}$.
3. $\mu = 100$, $V = 100$, $\delta = 30$: $P(|X - 100| \geq 30) \leq \frac{100}{900} = \frac{1}{9}$. D'après la question 2 avec $\delta = 30$, il faut $q\delta^2 = 100$, soit $q = \frac{1}{9}$: X prend les valeurs 70, 100, 130 avec les probabilités $\frac{1}{18}$, $\frac{8}{9}$, $\frac{1}{18}$. Alors $E(X) = 100$, $V(X) = \frac{1}{9} \times 900 = 100$ et $P(|X - 100| \geq 30) = \frac{1}{9}$: la borne est atteinte.
4. Pour tout couple (V, δ) avec $0 < V < \delta^2$, il existe une variable qui réalise l'égalité (question 2 avec $q = \frac{V}{\delta^2}$). On ne peut donc pas diminuer le majorant $\frac{V}{\delta^2}$ en ne connaissant que E et V . Pour une loi connue comme une binomiale, la valeur exacte est en revanche bien plus petite (D5), car on dispose de plus d'information.

Conclusion : l'inégalité de Bienaymé-Tchebychev est optimale parmi les majorations qui n'utilisent que l'espérance et la variance.

D8.

1. X_i vaut 1 si la i -ème épreuve est un succès. La somme S compte donc le nombre de succès lors de n épreuves de Bernoulli identiques et indépendantes : $S \sim \mathcal{B}(n, p)$.
 $E(X_1) = 1 \times p + 0 \times (1 - p) = p$; $E(X_1^2) = p$ car $X_1^2 = X_1$; $V(X_1) = p - p^2 = p(1 - p)$.
2. Par linéarité de l'espérance : $E(S) = E(X_1) + \dots + E(X_n) = np$. Les X_i sont indépendantes, donc les variances s'additionnent : $V(S) = V(X_1) + \dots + V(X_n) = np(1 - p)$.
3. Ici $E(S) = 200 \times 0,3 = 60$ et $V(S) = 200 \times 0,3 \times 0,7 = 42$. L'événement $40 < S < 80$ s'écrit $|S - 60| < 20$. Avec $\delta = 20$: $P(|S - 60| \geq 20) \leq \frac{42}{400} = 0,105$, donc $P(40 < S < 80) \geq 1 - 0,105 = 0,895$.

Conclusion : $E(S) = np$, $V(S) = np(1 - p)$; $P(40 < S < 80) \geq 0,895$ **pour** $\mathcal{B}(200; 0,3)$.

D9.

1. $E(X) = \frac{1 + 2 + \dots + 6}{6} = 3,5$ et $E(X^2) = \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6}$, donc $V(X) = \frac{91}{6} - \frac{49}{4} =$

$\frac{182 - 147}{12} = \frac{35}{12}$ et $\sigma = \sqrt{35/12} \approx 1,708$. Pour $n = 100$: $E(M_n) = 3,5$, $V(M_n) = \frac{35/12}{100} = \frac{7}{240} \approx 0,0292$ et $\sigma(M_n) = \frac{\sigma}{\sqrt{100}} \approx 0,171$.

2. $\bar{m}$ est la moyenne des 5 000 moyennes d'échantillons observées (proche de $E(M_n) = 3,5$) et $\bar{s}$ est leur écart type observé. On s'attend à $\bar{s}$ proche de $\sigma/\sqrt{n} \approx 0,171$ (le script affiche la valeur théorique 0,171 à côté). Pour $n = 400$, $\sigma/\sqrt{400} \approx 0,085$: multiplier la taille de l'échantillon par 4 divise l'écart type par 2.
3. Avec $\delta = k\sigma/\sqrt{n}$, l'inégalité de concentration donne $P(|M_n - \mu| \geq \delta) \leq \frac{\sigma^2}{n\delta^2} = \frac{1}{k^2}$, donc $P(|M_n - \mu| < \delta) \geq 1 - \frac{1}{k^2}$: minoration 0 pour $k = 1$ (aucune information), 0,75 pour $k = 2$ et $\frac{8}{9} \approx 0,889$ pour $k = 3$. Les proportions observées (environ 0,69, 0,96, 0,998) sont bien supérieures aux minoration garanties : l'inégalité est valable pour toute loi, donc loin d'être précise pour la moyenne de 100 dés.
4. Par exemple :

```
prop = sum(1 for x in moyennes if abs(x - mu) <= 0.3) / N
print(prop)
```

L'inégalité de concentration avec $\delta = 0,3$ donne $P(|M_n - \mu| \geq 0,3) \leq \frac{35/12}{100 \times 0,09} = \frac{35}{108} \approx 0,324$, donc $P(|M_n - \mu| < 0,3) \geq 0,676$. Le script affiche une proportion voisine de 0,93 (valeur exacte 0,926 à 10^{-3} près), bien supérieure à la minoration 0,676 : la minoration est correcte mais pessimiste.