

Sommes de variables aléatoires et concentration

Fiche Défi

Nom : _____

Date : _____

Exercice D1 : Démonstration de l'inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire prenant les valeurs distinctes $x_1, \dots, x_r$ avec les probabilités $p_1, \dots, p_r$. On note $\mu = E(X)$, $V = V(X)$, et δ un réel strictement positif. Soit I l'ensemble des indices i tels que $|x_i - \mu| \geq \delta$.

- Justifier que $V \geq \sum_{i \in I} p_i (x_i - \mu)^2 \geq \delta^2 \sum_{i \in I} p_i$. En déduire que $P(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$.
- On suppose $V > 0$ et on note $\sigma = \sqrt{V}$. Montrer que, pour tout réel $k > 0$, $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.
Donner les majorations obtenues pour $k = 2$ et $k = 3$.
- On suppose $V = 0$. Montrer que $P(|X - \mu| \geq \delta) = 0$ pour tout $\delta > 0$. En déduire que $P(X = \mu) = 1$.
(On pourra choisir δ égal au plus petit des écarts non nuls $|x_i - \mu|$.)

Exercice D2 : Indépendance et variance d'une somme

Soient X et Y deux variables aléatoires indépendantes, prenant les valeurs $x_1, \dots, x_r$ et $y_1, \dots, y_s$. On

admet que $E(XY) = \sum_{i=1}^r \sum_{j=1}^s x_i y_j P(X = x_i \text{ et } Y = y_j)$.

- Montrer que $E(XY) = E(X)E(Y)$.
- En développant $V(X+Y) = E((X+Y)^2) - (E(X)+E(Y))^2$, montrer que $V(X+Y) = V(X)+V(Y)$.
- On admet que X et $-Y$ sont indépendantes. Montrer que $V(X-Y) = V(X) + V(Y)$.
- Vérifier la formule de la question 2 pour deux variables de Bernoulli indépendantes de même paramètre p , en déterminant la loi de leur somme.
- Pour aller plus loin, hors programme.* La réciproque de la question 1 est-elle vraie ? Soit X uniforme sur $\{-2; -1; 1; 2\}$ et $Y = X^2$. Calculer $E(X)E(Y)$ et $E(XY)$, puis étudier l'indépendance de X et Y .

Exercice D3 : Les chapeaux du vestiaire

Soit $n \geq 2$. Dans un vestiaire, n personnes déposent leur chapeau, puis les chapeaux sont rendus au hasard : chacune des $n!$ façons de répartir les n chapeaux entre les n personnes est équiprobable. On note N le nombre de personnes qui retrouvent leur propre chapeau, et X_i la variable qui vaut 1 si la personne i retrouve son chapeau et 0 sinon.

- Montrer que $P(X_i = 1) = \frac{1}{n}$. En déduire $E(N)$: que remarque-t-on ?
- Pour $i \neq j$, montrer que $E(X_i X_j) = \frac{1}{n(n-1)}$. Comparer avec $E(X_i)E(X_j)$: peut-on appliquer l'additivité de la variance à $N = X_1 + \dots + X_n$?
- En développant $N^2 = \sum_i X_i^2 + \sum_{i \neq j} X_i X_j$, calculer $E(N^2)$, puis $V(N)$.
- Cas $n = 4$: montrer que $P(N = 3) = 0$, $P(N = 4) = \frac{1}{24}$, $P(N = 2) = \frac{1}{4}$, puis déterminer la loi de N et vérifier $E(N)$ et $V(N)$.
- Pour $n = 10$, majorer $P(N \geq 6)$.

Exercice D4 : Combiner deux mesures

Deux instruments mesurent la même grandeur m . Les mesures X et Y sont des variables aléatoires indépendantes, d'espérance m et de variances $V(X) = 4$ et $V(Y) = 1$. Pour $a \in [0; 1]$, on pose $Z_a = aX + (1-a)Y$.

- Calculer $E(Z_a)$. Exprimer $V(Z_a)$ en fonction de a .
- Déterminer la valeur de a qui minimise $V(Z_a)$ et la variance minimale.

3. Comparer cette variance minimale à $V(X)$ et à $V(Y)$. Que peut-on en conclure sur l'intérêt de combiner les deux mesures ?
4. Pour cette valeur optimale de a , minorer $P(|Z_a - m| < 2)$. Quelle minoration obtient-on avec X seul ?
5. Généralisation : $V(X) = \sigma_1^2$ et $V(Y) = \sigma_2^2$ avec $\sigma_1, \sigma_2 > 0$. Montrer que la variance minimale vaut $\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, atteinte pour $a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

Exercice D5 : La borne est-elle réaliste ?

On lance n fois une pièce équilibrée et on note F_n la fréquence de « pile ». On s'intéresse à l'événement $A_n = (|F_n - 0,5| \geq 0,05)$.

1. Exprimer avec l'inégalité de concentration un majorant b_n de $P(A_n)$ en fonction de n .
2. Soit X_n le nombre de « pile ». Montrer que $A_n = (10 |2X_n - n| \geq n)$.
3. Le script suivant calcule $P(A_n)$ (les entiers de Python sont exacts, la division finale est faite en flottant).

```
from math import comb

def risque(n):
    total = sum(comb(n, k) for k in range(n + 1)
                if 10 * abs(2 * k - n) >= n)
    return total / 2**n

for n in (100, 400, 1000, 2000):
    print(n, risque(n))
```

Expliquer pourquoi `risque(n)` vaut $P(A_n)$. Le script affiche 0,3682..., 0,0510..., 0,001730... et $8,457... \times 10^{-6}$. Comparer, pour chaque valeur de n , ces nombres à b_n .

4. Commenter : que révèle cette comparaison sur la vitesse de décroissance du risque exact et sur celle de b_n ? Pourquoi utilise-t-on tout de même l'inégalité de concentration ?

Exercice D6 : Retards de livraison (type bac)

Une entreprise emploie deux livreurs. Chaque colis du livreur A est livré en retard avec la probabilité 0,08, et chaque colis du livreur B avec la probabilité 0,05, les retards des colis étant indépendants. Un jour donné, A livre 50 colis et B en livre 80. On note N et M les nombres de colis en retard de A et de B , et $T = N + M$. On admet que N et M sont indépendantes.

Partie A : étude de T

1. Justifier que N et M suivent des lois binomiales dont on précisera les paramètres. Calculer $E(N)$, $V(N)$, $E(M)$ et $V(M)$.
2. Calculer $E(T)$, $V(T)$ et $\sigma(T)$ (arrondi au centième).
3. Montrer que T ne suit aucune loi binomiale. (On supposera que $T \sim \mathcal{B}(n, q)$ et on calculera q , puis n .)

Partie B : majorations

4. Montrer que $P(T \geq 20) \leq 0,052$.
5. Minorer la probabilité que T soit compris entre 3 et 13 (au sens large).

Partie C : estimer une proportion

6. L'entreprise veut estimer la proportion p de colis livrés en retard, connue seulement par $p \leq 0,1$. Montrer que $p(1 - p) \leq 0,09$ pour tout $p \in [0; 0,1]$.
7. On étudie n colis indépendants et on note F_n la fréquence de colis en retard. Déterminer le plus petit entier n garantissant $P(|F_n - p| < 0,01) \geq 0,95$ pour tout $p \in [0; 0,1]$. Quel serait n sans l'information $p \leq 0,1$?

Partie D : Python

8. Écrire une fonction `retards(n, p)` qui simule le nombre de retards parmi n colis, puis une fonction `estime(N)` qui renvoie la fréquence des jours où $T \geq 20$ sur N journées simulées. À quel ordre de grandeur s'attendre d'après la question 4 ?

Exercice D7 : L'inégalité de Bienaymé-Tchebychev est optimale

1. Soit X de loi : $P(X = -3) = 0,08$, $P(X = 0) = 0,84$, $P(X = 3) = 0,08$. Calculer $E(X)$, $V(X)$, puis comparer $P(|X| \geq 3)$ à la majoration de Bienaymé-Tchebychev pour $\delta = 3$.
2. Généralisation : soient $\mu \in \mathbb{R}$, $\delta > 0$ et $q \in]0; 1[$. Soit X prenant les valeurs $\mu - \delta$, μ , $\mu + \delta$ avec les probabilités $\frac{q}{2}$, $1 - q$, $\frac{q}{2}$. Montrer que $E(X) = \mu$, $V(X) = q\delta^2$ et que $P(|X - \mu| \geq \delta) = \frac{V(X)}{\delta^2}$.
3. Une variable a pour espérance 100 et écart-type 10. Majorer $P(|X - 100| \geq 30)$. Construire une variable de loi à trois valeurs qui atteint exactement cette borne.
4. Que peut-on conclure sur la possibilité d'améliorer l'inégalité sans hypothèse supplémentaire sur la loi ? Comparer avec le cas d'une loi binomiale (voir D5).

Exercice D8 : Espérance et variance de la loi binomiale

Soit $n \geq 1$ et $p \in]0; 1[$. On considère n variables aléatoires $X_1, \dots, X_n$ de Bernoulli de paramètre p , mutuellement indépendantes, et $S = X_1 + \dots + X_n$.

1. Justifier que S suit la loi $\mathcal{B}(n, p)$. Calculer $E(X_1)$ et $V(X_1)$.
2. En déduire $E(S) = np$ et $V(S) = np(1 - p)$. Préciser quelle propriété sert pour chacun des deux résultats.
3. Soit S de loi $\mathcal{B}(200; 0,3)$. Minorer $P(40 < S < 80)$.

Exercice D9 : Simuler des échantillons et comparer à la théorie

On lance $n = 100$ fois un dé équilibré à six faces et on note M_n la moyenne des n résultats. On simule $N = 5\,000$ échantillons de taille n .

```
from random import randint
from math import sqrt

n, N = 100, 5000
mu, sigma = 3.5, sqrt(35 / 12)

moyennes = []
for j in range(N):
    echantillon = [randint(1, 6) for i in range(n)]
    moyennes.append(sum(echantillon) / n)

m = sum(moyennes) / N
s = sqrt(sum((x - m)**2 for x in moyennes) / N)
print(round(s, 3), round(sigma / sqrt(n), 3))
for k in (1, 2, 3):
    prop = sum(1 for x in moyennes
               if abs(x - mu) <= k * sigma / sqrt(n)) / N
    print(k, prop)
```

1. Soit X le résultat d'un lancer. Calculer $E(X)$ et $V(X)$ (on retrouvera $\sigma^2 = \frac{35}{12}$), puis $E(M_n)$, $V(M_n)$ et $\sigma(M_n)$ pour $n = 100$.
2. Que représentent les variables `m` et `s` ? Quelles valeurs approximatives s'attend-on à lire pour `s` ? Que devient cette valeur si l'on prend $n = 400$?
3. Pour $k = 1, 2, 3$, l'inégalité de concentration fournit une minoration de $P(|M_n - \mu| \leq k\sigma/\sqrt{n})$. La

déterminer. Le script affiche les proportions 0,69, 0,96 et 0,998 environ : les commenter.

4. Écrire une version du script qui mesure la proportion d'échantillons pour lesquels $|M_n - \mu| \leq 0,3$.
Calculer par l'inégalité de concentration la minoration de la probabilité correspondante et comparer.