

Approfondissement

Concentration et loi des grands nombres

Pour aller plus loin : démonstration de Bienaymé-Tchebychev, inégalité de Markov, cas d'égalité, théorème de Bernoulli, marche aléatoire en $\sqrt{n}$

6 exercices · niveau ●●● · durée conseillée : 2 h 30 · calculatrice autorisée

Exercice 1 Démontrer l'inégalité de Bienaymé-Tchebychev

●●●

Une variable aléatoire X prend les valeurs $x_1, \dots, x_r$ avec les probabilités $p_1, \dots, p_r$; on note $\mu = E(X)$ et $V = V(X)$. Soit $\varepsilon > 0$. On note I l'ensemble des indices i de $\{1; \dots; r\}$ tels que $|x_i - \mu| \geq \varepsilon$, et J l'ensemble des autres indices.

1.1 Justifier que $V = \sum_{i=1}^r p_i(x_i - \mu)^2$ et que $P(|X - \mu| \geq \varepsilon) = \sum_{i \in I} p_i$.

1.2 Justifier que $\sum_{i \in J} p_i(x_i - \mu)^2 \geq 0$, puis en déduire que $\sum_{i \in I} p_i(x_i - \mu)^2 \leq V$.

1.3 Démontrer que, pour tout $i \in I$, $p_i(x_i - \mu)^2 \geq \varepsilon^2 p_i$, puis que $\varepsilon^2 \sum_{i \in I} p_i \leq V$.

1.4 Conclure : $P(|X - \mu| \geq \varepsilon) \leq \frac{V}{\varepsilon^2}$. Où l'hypothèse $\varepsilon > 0$ a-t-elle été utilisée ?

Exercice 2 L'inégalité de Markov (hors programme — pour aller plus loin)

●●●

Soit Y une variable aléatoire prenant les valeurs $y_1, \dots, y_r$, **toutes positives ou nulles**, avec les probabilités $q_1, \dots, q_r$, et soit $a > 0$.

2.1 En s'inspirant de l'exercice précédent, démontrer l'*inégalité de Markov* : $P(Y \geq a) \leq \frac{E(Y)}{a}$.

2.2 Application : dans une entreprise, le salaire mensuel moyen est de 2 400 €. Majorer la proportion de salariés gagnant au moins 7 200 € par mois. Cette majoration dépend-elle de la répartition des salaires ?

2.3 Soit X une variable aléatoire d'espérance μ et de variance V , et $\varepsilon > 0$. En appliquant l'inégalité de Markov à $Y = (X - \mu)^2$ et à $a = \varepsilon^2$, retrouver l'inégalité de Bienaymé-Tchebychev.

Exercice 3 L'inégalité peut-elle être améliorée ?

●●●

Soient μ un réel, $\varepsilon > 0$ et $t \in]0; 1]$. On considère la variable aléatoire X prenant les valeurs $\mu - \varepsilon$, μ et $\mu + \varepsilon$ avec les probabilités respectives $\frac{t}{2}$, $1 - t$ et $\frac{t}{2}$.

3.1 Vérifier que ces trois probabilités définissent bien une loi, puis démontrer que $E(X) = \mu$ et $V(X) = t\varepsilon^2$.

3.2 Calculer $P(|X - \mu| \geq \varepsilon)$ et comparer avec le majorant $\frac{V(X)}{\varepsilon^2}$: qu'en conclut-on sur l'inégalité de Bienaymé-Tchebychev ? Traiter l'exemple $\mu = 10$, $\varepsilon = 2$, majorant égal à 0,25 : préciser alors la loi de X et $V(X)$.

3.3 Une variable aléatoire B suit une loi de Bernoulli de paramètre p . Étudier la fonction f définie sur $[0; 1]$ par $f(p) = p(1 - p)$ et en déduire la variance maximale d'une loi de Bernoulli. Pourquoi cette majoration est-elle utile dans l'étude d'un sondage ?

Exercice 4 De Bienaymé-Tchebychev au théorème de Bernoulli



Soit X une variable aléatoire d'espérance μ et de variance V , et soit $(X_1; \dots; X_n)$ un échantillon de taille n de la loi de X . On pose $S_n = X_1 + \dots + X_n$ et $M_n = \frac{S_n}{n}$.

- 4.1 En utilisant les résultats du chapitre sur les sommes de variables aléatoires, démontrer que $E(M_n) = \mu$ et $V(M_n) = \frac{V}{n}$.
- 4.2 En appliquant l'inégalité de Bienaymé-Tchebychev à M_n , démontrer l'inégalité de concentration, puis en déduire la loi des grands nombres.
- 4.3 On suppose maintenant que X suit une loi de Bernoulli de paramètre p : M_n est alors la fréquence F_n du succès dans l'échantillon. Démontrer le *théorème de Bernoulli* : pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} P(|F_n - p| \geq \varepsilon) = 0$.
- 4.4 Déterminer le plus petit entier n tel que l'inégalité obtenue garantisse $P(|F_n - p| \geq 0,01) \leq 0,01$, quelle que soit la valeur de p .

Exercice 5 L'ordre de grandeur d'une marche aléatoire



Un pion part de l'origine et, à chaque étape, avance de 1 ou recule de 1 avec la probabilité 0,5, indépendamment des étapes précédentes. On note S_n son abscisse après n étapes : $E(S_n) = 0$ et $V(S_n) = n$.

- 5.1 Soit $\lambda > 0$. Démontrer que, pour tout entier $n \geq 1$, $P(|S_n| \geq \lambda\sqrt{n}) \leq \frac{1}{\lambda^2}$. Ce majorant dépend-il de n ?
- 5.2 En déduire que la probabilité que le pion se trouve, après n étapes, à une distance de l'origine supérieure à $3\sqrt{n}$ n'excède jamais $\frac{1}{9}$. Application numérique pour $n = 10\,000$.
- 5.3 Démontrer que, pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} P(|S_n| \geq \varepsilon n) = 0$. Interpréter : le pion s'éloigne de l'origine, mais beaucoup plus lentement que le nombre d'étapes.
- 5.4 Recopier et compléter les lignes 5 et 13 du programme ci-contre, qui estime sur N marches la proportion de celles vérifiant $|S_n| \geq \lambda\sqrt{n}$. Pour $\lambda = 2$, la valeur renvoyée par `proportion(10000, 400, 2)` est d'environ 0,050 : comparer au majorant de la question 5.1 et commenter.

```

1 from random import randint
2 from math import sqrt
3
4 def marche(n):
5     s = ...
6     for i in range(n):
7         s = s + 2*randint(0, 1) - 1
8     return s
9
10 def proportion(N, n, lam):
11     c = 0
12     for j in range(N):
13         if ... >= lam*sqrt(n):
14             c = c + 1
15     return c / N

```

Exercice 6 Pourquoi les instituts interrogent-ils mille personnes ?



Les sondages publiés dans la presse reposent très souvent sur un échantillon de 1 000 personnes et sont annoncés avec une « marge d'erreur de 3 points ». On note p la proportion inconnue d'un caractère dans la population et F_n la fréquence observée sur un échantillon de taille n .

- 6.1 Déterminer le plus petit entier n pour lequel l'inégalité de concentration garantit $P(|F_n - p| \geq 0,03) \leq 0,05$ sans hypothèse sur p .
- 6.2 Avec $n = 1\,000$ et un risque d'au plus 0,05, quelle précision ε l'inégalité de concentration permet-elle de garantir ? Donner la valeur exacte, puis l'arrondi à 10^{-3} .
- 6.3 On suppose maintenant $p = 0,5$ et $n = 1\,000$. Justifier que 1 000 F_{1000} suit une loi binomiale, puis calculer $P(|F_{1000} - 0,5| \geq 0,03)$ à la calculatrice (arrondi à 10^{-3}).
- 6.4 Comparer les trois réponses précédentes et expliquer pourquoi la pratique des instituts ne contredit pas le cours. (*Hors programme : les instituts approchent la loi binomiale par la loi normale, ce qui donne une approximation bien plus fine.*)