

Fiche de cours

Concentration et loi des grands nombres

L'essentiel du chapitre en deux pages

définitions · propriétés · méthodes · à savoir par cœur

1 MESURER UN ÉCART À L'ESPÉRANCE

Notation. Pour une variable aléatoire X d'espérance μ , l'événement « X s'écarte de μ d'au moins ε » s'écrit $|X - \mu| \geq \varepsilon$, c'est-à-dire $X \leq \mu - \varepsilon$ ou $X \geq \mu + \varepsilon$. Son contraire est $\mu - \varepsilon < X < \mu + \varepsilon$.

Théorème — inégalité de Bienaymé-Tchebychev. Si X a pour espérance μ et pour variance V , alors pour tout réel $\varepsilon > 0$:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{V}{\varepsilon^2}, \quad \text{donc} \quad P(\mu - \varepsilon < X < \mu + \varepsilon) \geq 1 - \frac{V}{\varepsilon^2}.$$

Propriété — écarts en multiples de σ . Si $\sigma > 0$, en prenant $\varepsilon = k\sigma$ avec $k > 0$: $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$. Ainsi $P(|X - \mu| \geq 2\sigma) \leq 0,25$ et $P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9} \approx 0,11$: l'écart type est l'**unité naturelle de dispersion**.

Remarque. Cette majoration est **universelle** : elle ne suppose rien sur la loi de X . En contrepartie elle est souvent très large, et n'a d'intérêt que si $\frac{V}{\varepsilon^2} < 1$.

2 MOYENNE D'UN ÉCHANTILLON

Définition. Un **échantillon de taille n** de la loi de X est une liste $(X_1, \dots, X_n)$ de variables aléatoires indépendantes suivant toutes cette loi. On pose $S_n = X_1 + \dots + X_n$ et $M_n = \frac{S_n}{n}$.

Propriété. Si $E(X) = \mu$, $V(X) = V$ et $\sigma(X) = \sigma$, alors

$$E(S_n) = n\mu, \quad V(S_n) = nV; \quad E(M_n) = \mu, \quad V(M_n) = \frac{V}{n}, \quad \sigma(M_n) = \frac{\sigma}{\sqrt{n}}.$$

3 INÉGALITÉ DE CONCENTRATION ET LOI DES GRANDS NOMBRES

Théorème — inégalité de concentration. En appliquant Bienaymé-Tchebychev à M_n , pour tout réel $\varepsilon > 0$:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{V}{n\varepsilon^2}.$$

Théorème — loi des grands nombres. Pour tout réel $\varepsilon > 0$: $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \varepsilon) = 0$.

Remarque. Autrement dit, la probabilité que la moyenne observée sur un échantillon s'écarte de μ d'au moins ε devient aussi petite que l'on veut pourvu que n soit assez grand : c'est la justification théorique de l'estimation d'une espérance (ou d'une proportion) par une moyenne observée.

4 DÉTERMINER UNE TAILLE D'ÉCHANTILLON

Méthode — précision ε et risque α . Pour garantir $P(|M_n - \mu| \geq \varepsilon) \leq \alpha$, il suffit d'imposer $\frac{V}{n\varepsilon^2} \leq \alpha$, c'est-à-dire

$$n \geq \frac{V}{\alpha\varepsilon^2}.$$

On prend alors pour n le plus petit entier vérifiant cette inégalité.

Remarque. Cas d'une **proportion** : si X suit la loi de Bernoulli de paramètre p , alors $V = p(1 - p) \leq \frac{1}{4}$ pour tout p de $[0; 1]$. On peut donc toujours majorer V par $\frac{1}{4}$, ce qui donne une taille $n \geq \frac{1}{4\alpha\varepsilon^2}$ valable quelle que soit la valeur inconnue de p .

Méthode — comparer à la probabilité exacte. Lorsque la loi de X est connue (souvent binomiale), calculer la probabilité exacte à la calculatrice et la comparer à la majoration : on constate que Bienaymé-Tchebychev est correcte mais peu précise.

Remarque — simulation. On illustre ces résultats en simulant N échantillons de taille n et en comptant la proportion de moyennes M_n situées à moins de $k \frac{\sigma}{\sqrt{n}}$ de μ : la probabilité correspondante est au moins égale au minorant $1 - \frac{1}{k^2}$, et la proportion observée le dépasse en pratique nettement (environ 0,95 pour $k = 2$, contre 0,75) ; elle varie peu avec n ; pour un écart ε fixé, en revanche, la proportion de moyennes à moins de ε de μ augmente avec n : l'histogramme des M_n se resserre autour de μ .

5 SAVOIR-FAIRE DU CHAPITRE

► SAVOIR-FAIRE

- Traduire un écart à l'espérance par une valeur absolue et passer à l'événement contraire.
- Majorer $P(|X - \mu| \geq \varepsilon)$ par l'inégalité de Bienaymé-Tchebychev, en multiples de σ si besoin.
- Calculer $E(M_n)$, $V(M_n)$, $\sigma(M_n)$ puis appliquer l'inégalité de concentration.
- Déterminer une taille d'échantillon assurant une précision ε avec un risque au plus α , en majorant $p(1-p)$ par $\frac{1}{4}$ pour une proportion.
- Énoncer et interpréter la loi des grands nombres ; comparer la majoration obtenue à la probabilité exacte.
- Interpréter une simulation de N échantillons de taille n à la lumière de la concentration de M_n .

6 EXEMPLES TYPES

★ EXEMPLE — majorer une probabilité

On lance 100 fois une pièce équilibrée ; X , nombre de « pile », suit $\mathcal{B}(100; 0,5)$, donc $\mu = 50$ et $V = 25$. Avec $\varepsilon = 10$: $P(|X - 50| \geq 10) \leq \frac{25}{100} = 0,25$, donc $P(40 < X < 60) \geq 0,75$. La valeur exacte de $P(|X - 50| \geq 10)$, à la calculatrice, vaut environ 0,057 : la majoration est correcte mais large.

★ EXEMPLE — taille d'un échantillon

On veut estimer une proportion inconnue p à $\varepsilon = 0,02$ près avec un risque $\alpha = 0,05$. En majorant $V = p(1-p)$ par $\frac{1}{4}$: $n \geq \frac{1}{4 \times 0,05 \times 0,02^2} = 12\,500$. Un échantillon de 12 500 personnes suffit pour garantir $P(|M_n - p| \geq 0,02) \leq 0,05$.

★ EXEMPLE — écart en multiples de σ

Une pièce usinée a une longueur X d'espérance $\mu = 40$ mm et d'écart type $\sigma = 0,3$ mm. Alors $P(|X - 40| \geq 0,9) = P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9}$: au plus une pièce sur neuf (environ 11,1 %) s'écarte d'au moins 0,9 mm de la longueur moyenne.

★ EXEMPLE — inégalité de concentration

On lance 1 000 fois un dé équilibré et l'on note M_{1000} la moyenne des résultats obtenus. Ici $\mu = 3,5$ et $V = \frac{35}{12} \approx 2,92$, donc $P(|M_{1000} - 3,5| \geq 0,2) \leq \frac{35/12}{1\,000 \times 0,2^2} \approx 0,073$: avec une probabilité d'au moins 0,927, la moyenne observée est strictement comprise entre 3,3 et 3,7.

✗ PIÈGES

- Bienaymé-Tchebychev donne une **majoration**, jamais une valeur : écrire $P(|X - \mu| \geq \varepsilon) = \frac{V}{\varepsilon^2}$ est faux.
- Ne pas confondre V et σ : le dénominateur contient ε^2 et le numérateur la **variance**, pas l'écart type.
- Dans l'inégalité de concentration, c'est la variance de la loi initiale qui apparaît, divisée par n : $\frac{V}{n\varepsilon^2}$, et non $\frac{V}{\varepsilon^2}$.
- Une majoration supérieure ou égale à 1 n'apporte aucune information : l'inégalité n'est utile que pour ε assez grand, ou n assez grand.
- La loi des grands nombres porte sur la **probabilité** d'un écart, pas sur une compensation des résultats passés : un tirage n'est jamais « en retard ».
- Multiplier la précision par 2 (c'est-à-dire diviser ε par 2) multiplie la taille d'échantillon nécessaire par 4, et non par 2.