Chapitre 15 – Cours

Concentration et loi des grands nombres

Dans ce chapitre, tu vas apprendre à :

- Appliquer une inégalité probabiliste
- Interpréter la convergence de fréquences

Comprendre les notions

Une borne générale

Définition – Notion de départ

Pour une variable d'espérance m et de variance finie σ^2, l'inégalité de Bienaymé-Tchebychev donne, pour tout $a > 0$, la borne suivante :

$$P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Cette borne ne donne pas en général la valeur exacte de la probabilité. Si le quotient dépasse un, la borne triviale un est meilleure. Le seuil mesure un écart absolu à l'espérance ; il faut l'exprimer dans la même unité que la variable.

Moyennes et échantillonnage

Propriété – Lecture et conséquences

Pour la moyenne de n observations indépendantes de même loi, la variance vaut σ^2/n. L'inégalité devient

$$P(|M_n - m| \geq a) \leq \frac{\sigma^2}{na^2}.$$

À seuil fixé, cette borne tend vers zéro lorsque n augmente. C'est une forme de loi des grands nombres : la moyenne se concentre en probabilité autour de m. Cela ne signifie pas que toute nouvelle observation est plus proche de m, ni que les fréquences fluctuent de façon monotone.

Première lecture : application et dimensionnement

Propriété – Lecture et conséquences

Pour une fréquence de succès d'épreuves de Bernoulli, m=p et $\sigma^2 = p(1-p) \leq 1/4$. Donc $P(|F_n - p| \geq a) \leq 1/(4na^2)$. Si l'on veut garantir une erreur au moins un dixième avec probabilité au plus un centième, une taille de deux mille cinq cents suffit avec cette borne uniforme. La taille obtenue est suffisante et souvent prudente ; elle n'est pas nécessairement optimale. Une simulation illustre le phénomène mais la garantie découle ici du raisonnement probabiliste.

Convergence en probabilité

Définition – Convergence en probabilité

Dire que la moyenne converge en probabilité vers m signifie que, pour tout écart fixé strictement positif, la probabilité de s'en éloigner davantage tend vers zéro. On n'affirme ni une égalité exacte à partir d'un certain rang ni une amélioration à chaque nouveau tirage.

$$\forall a > 0, \qquad P(|M_n - m| \geq a) \longrightarrow 0.$$

Dans le cadre étudié, l'indépendance, une même loi et une variance finie permettent d'obtenir cette conclusion par Bienaymé-Tchebychev.

Deux exemples résolus

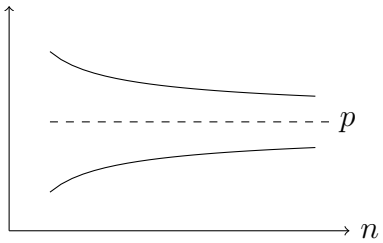
Exemple résolu – Contrôler une moyenne

1. Chaque observation a une espérance de cinquante et une variance de neuf. Pour cent observations indépendantes de même loi, la moyenne a une variance $9/100$.
2. Avec un seuil de 0,6, l'inégalité donne $P(|M_{100} - 50| \geq 0{,}6) \leq 0{,}09/0{,}36 = 0{,}25$.
3. Par complémentaire, $P(|M_{100} - 50| < 0{,}6) \geq 0{,}75$. Cette garantie ne prétend pas que la probabilité exacte vaut trois quarts.

Exemple résolu – Dimensionner pour une fréquence

1. On veut garantir $P(|F_n - p| \geq 0{,}05) \leq 0{,}05$, sans connaître p. On utilise $p(1-p) \leq 1/4$.
2. Il suffit de demander $1/(4n \times 0{,}05^2) \leq 0{,}05$. Comme $4 \times 0{,}05^2 = 0{,}01$, cette condition devient $n \geq 2000$.
3. Deux mille essais indépendants suffisent donc avec cette méthode. Une étude plus précise de la loi peut demander moins d'essais ; la borne n'affirme aucune nécessité.

Une représentation pour comprendre



Une bande d'échelle $1/\sqrt{n}$ se resserre autour de p ; elle n'est pas une trajectoire.

Les courbes sont les deux bornes déterministes $p \pm 1/\sqrt{n}$ entourant la valeur centrale. Elles illustrent une échelle de concentration, et non des trajectoires observées ou une garantie que chaque fréquence reste dans la bande. Les fluctuations d'une fréquence réelle peuvent encore augmenter entre deux tailles voisines. Le paramètre a mesure la précision, tandis que la probabilité maximale d'écart mesure la fiabilité. Diviser a par deux exige quatre fois plus d'observations si l'on conserve la même borne probabiliste.

Cette dépendance en carré explique le coût élevé d'une précision plus fine.

À retenir – Mini-bilan

Une inégalité fournit une garantie, pas une probabilité exacte. Choisir seuil et niveau d'erreur séparément ; résoudre l'inégalité en n, puis arrondir à un entier suffisant. La concentration porte sur la moyenne, pas sur chaque observation.