Concentration et loi des grands nombres

*Rappel : On considère toujours des variables aléatoires associées à des univers **finis**.*

1 Inégalité de Bienaymé-Tchebychev

L'idée est la suivante : la probabilité qu'une variable aléatoire prenne une certaine valeur est d'autant plus faible que cette valeur est plus éloignée de l'espérance. Par exemple, si la moyenne théorique des devoirs de philosophie d'une classe est 12, il est moins probable pour un élève d'obtenir 18 que d'obtenir 14.

Mais d'autre part, la variance (ou l'écart-type) mesure la dispersion (théorique) des valeurs. Plus la variance est petite, plus les valeurs sont concentrées autour de l'espérance. Il doit donc être plus difficile de s'éloigner de celle-ci, et donc la probabilité est là aussi plus faible.

Autrement dit, la probabilité que la v.a. considérée prenne une certaine valeur va être d'autant plus faible que cette valeur est plus éloignée de l'espérance ***et*** que la variance est plus petite. L'inégalité de Bienaymé-Tchebychev, qui met en forme cette idée, formule un lien entre la probabilité d'une valeur et sa distance par rapport à l'espérance en fonction de deux paramètres : l'espérance et la variance (ou l'écart-type) de la variable.

Théorème 1
Soit X une variable aléatoire d'espérance μ et de variance $\mathcal{V}$. Pour tout réel strictement positif δ :

$$P\big(|X-\mu| \geq \delta\big) \leq \frac{\mathcal{V}}{\delta^2}$$

Démonstration :
Soient x_1, x_2, ... x_n les valeurs prises par la variable aléatoire X. Posons, pour simplifier l'écriture : $P(X = x_i) = p_i$.

On écrit la formule de la variance : $\mathcal{V} = \sum_{i=1}^{n} p_i \times (x_i - \mu)^2$

Répartissons les valeurs x_i en deux groupes :

- G_1, contenant les valeurs qui sont à une distance de moins de δ de l'espérance μ : $|x_i - \mu| < \delta$
- G_2, contenant les valeurs qui sont à une distance d'au moins δ de l'espérance μ : $|x_i - \mu| \geq \delta$

Ainsi : $\mathcal{V} = \sum_{x_i \in G_1} p_i \times (x_i-\mu)^2 + \sum_{x_i \in G_2} p_i \times (x_i-\mu)^2$

Négligeons les valeurs appartenant au premier groupe. Comme les termes sont tous positifs, on obtient :
$\mathcal{V} \geq \sum_{x_i \in G_2} p_i \times (x_i-\mu)^2$.

Dans cette somme, toutes les valeurs sont dans G_2, donc vérifient : $|x_i - \mu| \geq \delta$; donc $(x_i-\mu)^2 \geq \delta^2$.

Par conséquent : $\mathcal{V} \geq \sum_{x_i \in G_2} p_i \times \delta^2$. On factorise : $\mathcal{V} \geq \delta^2 \sum_{x_i \in G_2} p_i$.

Il reste à réécrire $\sum_{x_i \in G_2} p_i$: c'est la somme des probabilités de toutes les valeurs x_i vérifiant $|x_i - \mu| \geq \delta$.

Autrement dit, c'est la probabilité de l'événement : $|X-\mu| \geq \delta$. Ainsi : $\mathcal{V} \geq \delta^2 P\big(|X-\mu| \geq \delta\big)$.

On obtient enfin en divisant par δ^2 : $P\big(|X-\mu| \geq \delta\big) \leq \frac{\mathcal{V}}{\delta^2}$. ■

Remarques :

1. On constate que la probabilité que X prenne une valeur varie en sens inverse de la distance entre cette valeur et l'espérance. C'est conforme à l'idée intuitive exposée plus haut.
2. Pour illustrer simplement ce résultat, posons : $\delta = k\sigma$, où σ est l'écart-type de X et k un entier strictement positif. Alors : $\frac{\mathcal{V}}{\delta^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$.

La probabilité que X prenne une valeur distante de μ de plus de k écarts-type est inférieure à $\dfrac{1}{k^2}$.

$P(|X-\mu| \geq \sigma) \leq 1$ (pas très intéressant, je vous l'accorde) ;

$P(|X-\mu| \geq 2\sigma) \leq \dfrac{1}{4}$;

$P(|X-\mu| \geq 3\sigma) \leq \dfrac{1}{9}$ etc.

3. le caractère universel de l'inégalité de Bienaymé-Tchebychev a pour contrepartie le fait qu'elle est loin d'être optimale : ainsi, elle montre qu'un écart à μ supérieur à 2σ est de probabilité inférieure ou égale à $1/4$ alors que cette probabilité est souvent majorée par 0,05.C'est l'objet de l'exemple suivant.

Exemple : On effectue des série de 100 lancers d'une pièce ; la v.a. X qui compte le nombre de « pile » suit la loi binomiale de paramètres $n = 100$ et $p = 0{,}5$.

D'après le cours, on sait que $\mu = E(X) = np = 50$ et $\sigma = \sigma(X) = \sqrt{np(1-p)} = \sqrt{25} = 5$.

On trouve grâce à python ou à la calculatrice :

$P(|X-50| \geq 2\times 5) = P(X \leq 40 \cup X \geq 60) = 1 - P(41 \leq X \leq 59) \simeq 0{,}0569$ nettement inférieur à $\dfrac{1}{4}$.

$P(|X-50| \geq 3\times 5) = P(X \leq 35 \cup X \geq 65) = 1 - P(36 \leq X \leq 64) \simeq 0{,}0035$ nettement inférieur à $\dfrac{1}{9}$.

2 Inégalité de concentration et loi des grands nombres

Soit M_n la variable aléatoire moyenne d'un échantillon $(X_1 ; X_2 ; \dots ; X_n)$ de taille n d'une loi de probabilité d'espérance μ et de variance $\mathcal{V}$.

D'après le cours sur la somme de variables aléatoires, on sait que : $E(M_n) = \mu$ et $V(M_n) = \dfrac{\mathcal{V}}{n}$.

Appliquons l'inégalité de Bienaymé-Tchebychev à M_n :

Propriété 1 (Inégalité de concentration)
Pour tout réel $\delta > 0$, $P\big(|M_n - \mu| \geq \delta\big) \leq \dfrac{\mathcal{V}}{n\delta^2}$.

Ce qu'il faut surtout remarquer, c'est la présence du n au dénominateur de la fraction au second membre. Ceci implique que, pour tout $\delta > 0$ fixé, la probabilité que la v.a. moyenne de l'échantillon prenne une valeur distante de μ de plus de δ tend vers 0 lorsque n tend vers $+\infty$.

En effet, on a : $0 \leq P\big(|M_n - \mu| \geq \delta\big) \leq \dfrac{\mathcal{V}}{n\delta^2}$.

Or, $\lim\limits_{n\to+\infty} \dfrac{\mathcal{V}}{n\delta^2} = 0$. D'après le théorème des gendarmes, on en déduit : $\lim\limits_{n\to+\infty} P\big(|M_n - \mu| \geq \delta\big) = 0$.

Théorème 2 (Loi faible des grands nombres)
Soit M_n la v.a. moyenne d'un échantillon de taille n d'une loi de probabilité d'espérance μ.
Pour tout réel $\delta > 0$: $\lim\limits_{n\to+\infty} P\big(|M_n - \mu| \geq \delta\big) = 0$

Dit d'une autre manière : $\lim\limits_{n\to+\infty} P\big(\mu - \delta < M_n < \mu + \delta\big) = 1$.

Exemple : On lance n fois une pièce et on détermine la fréquence d'apparition de « pile ». La probabilité que cette fréquence n'appartienne pas à l'intervalle $[0{,}499999 ; 0{,}500001]$ tend vers 0 quand n tend vers $+\infty$.

Autrement dit, cela devient infiniment peu probable...