Concentration, loi des grands nombres : approfondissement

Une élection en approche

Lors d'une élection, deux candidats s'opposent. L'idée de cet exercice est d'estimer certains résultats en interrogeant un échantillon de personnes.

Exercice n°1

Une élection oppose un candidat A à un candidat B.
Soit p la proportion d'électeurs dans la population totale, décidant à voter pour le candidat A.
On souhaite estimer cette proportion p.
On effectue pour cela un sondage auprès de n personnes. On suppose que chaque personne interrogée donne son intention réelle de vote. La population est suffisamment grande pour cela soit assimilé à un tirage avec remise.
On note X_i la variable aléatoire qui vaut 1 si le personne i interrogée vote pour le candidate A. La variable aléatoire X vaut 0 sinon.
On note enfin la moyenne de cet échantillon $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$.

1. Montrer que pour tout p dans $[0\,;\,1]$ on a $p(1-p) \leqslant \dfrac{1}{4}$.
2. Quelle est la loi suivie par chacun des X_i ? Donner leur espérance et leur variance.
3. Montrer que pour tout réel $\delta > 0$, $P(M_n - \delta < p < M_n + \delta) \geqslant 1 - \dfrac{1}{4n\delta^2}$.

   Si f est la valeur prise par M_n lors du sondage, on dit que $]f-\delta\,;\,f+\delta[$ est un intervalle de confiance pour p au niveau de confiance supérieur ou égal à $1 - \dfrac{1}{4x\delta^2}$.
4. Le sondage auprès de 1 000 personnes donne une fréquence de votants pour le candidat A égale à 0,55.
   Un intervalle de confiance pour p est alors $]0,55-\delta\,;\,0,55+\delta[$. On souhaite que cet intervalle se trouve à un niveau supérieur ou égal à 0,95.
   Montrer qu'il suffit pour cela que $\delta \geqslant 0,0707$.
5. On prend $\delta = 0,071$. Donner alors l'intervalle de confiance de p au seuil de 0,95. Peut-on affirmer que p est strictement supérieur à 50% avec un niveau de confiance d'au moins 95% ?
6. Le candidat A souhaite que l'amplitude de l'intervalle de confiance au seuil de 0,95 soit de 4% maximum. Combien de personnes doit-on interroger au minimum ?

> Correction des exercices

Exercice n°1

1. On considère la fonction $f : p \mapsto p(1-p)$.
   On a $f(p) = -p^2 + p$ qui est un polynôme du second degré. Il atteint son maximum en $p = -\dfrac{1}{-2} = 0,5$ et la valeur de ce maximum est $f(0,5) = 0,25$.
   Ainsi, pour tout p réel et en particulier pour $p \in [0\,;\,1]$, on a bien $p(1-p) \leqslant 0,25$.

2. Chacun des X_i suit la loi de Bernoulli de paramètre p.
   $\mu = p \times 1 + (1-p) \times 0 = p$.
   Puis $V(X) = p \times 1^2 + (1-p) \times 0^2 - \mu^2 = p(1-p)$.

3. Puisque le choix de chacune des personnes est indépendant, la liste $(X_1,\, X_2,\, ...,\, X_n)$ constitue un échantillon de taille n de X.
   M_n est donc la variable aléatoire moyenne de l'échantillon. En utilisant l'inégalité de concentration, on obtient :
   $$P(|M_n - p| \geqslant \delta) \leqslant \frac{V(X)}{n\delta^2}$$
   $$\Leftrightarrow \quad P(|M_n - p| \geqslant \delta) \leqslant \frac{p(1-p)}{n\delta^2}$$
   $$\Leftrightarrow \quad P(|M_n - p| \geqslant \delta) \leqslant \frac{1}{4n\delta^2}$$
   On a donc la probabilité de l'évènement contraire :
   $$P(M_n - \delta < p < M_n + \delta) \geqslant 1 - \frac{1}{4n\delta^2}$$

4. Le sondage donne $f = 0,55$.
   On veut que l'intervalle de confiance $]0,55 - \delta\,;\,0,55 + \delta[$ soit à un niveau de confiance d'au moins 95%.

   Il suffit donc que $1 - \dfrac{1}{4n\delta^2} \geqslant 0,95$. Autrement dit :
   $$1 - \frac{1}{4 \times 1000 \times \delta^2} \geqslant 0,95$$
   $$\delta^2 \geqslant 0,005$$
   Puisque $\delta > 0$ on ne prend que la racine positive soit $\delta = \sqrt{0,005} \approx 0,0707$.

5. L'intervalle de confiance de p devient $]0,55 - 0,071\,;\,0,55 + 0,071[$ soit]0,479 ; 0,621[.
   Dans ce cas, 50% est dans l'intervalle de confiance et on peut avoir en particulier une valeur de p inférieure à 50%.
   On ne peut donc pas affirmer que p est strictement supérieur à 50% au seuil de 95%.

6. On souhaite que $\delta \geqslant 0,02$.

   Or, puisque $1 - \dfrac{1}{4n\delta^2} \geqslant 0,95$ cela revient à :
   $$0,05 \geqslant \frac{1}{4n\delta^2}$$

$$\Leftrightarrow \quad n \geqslant \frac{1}{4 \times 0,05 \times \delta^2}$$

$$\Leftrightarrow \quad n \geqslant \frac{1}{4 \times 0,05 \times 0,02^2}$$

Ce qui nous donne $n \geqslant 12\ 500$.

On doit donc interroger au moins 12 500 personnes.