Ex - Estimation par inégalité gl15

Une population de personnes doit voter à l'élection présidentielle. La proportion souhaitant voter pour la candidate Mme A. est inconnue, on la note p. Pour approcher cette valeur, on effectue un sondage sur n personnes : l'échantillon est modélisé par une suite de variables aléatoires indépendantes $(X_1, \ldots, X_n)$ suivant chacune une loi de Bernoulli de paramètre p ($X_i = 1$ si le i-ème individu souhaite voter pour Mme A., $X_i = 0$ sinon).
On note $S_n = \sum_{i=1}^{n} X_i$ de sorte que $\frac{S_n}{n}$ représente la proportion de personnes votant pour Mme A. dans l'échantillon.

1. Quelle est la loi suivie par S_n ?

✓ S_n suit une loi binomiale $\mathcal{B}(n, p)$.

2. Déterminer l'espérance et la variance de $\frac{S_n}{n}$.

✓ On en déduit que $\mathbb{E}(S_n) = np$ et $V(S_n) = np(1-p)$. D'après les propriétés de l'espérance et de la variance, on en déduit que $\mathbb{E}\left(\frac{S_n}{n}\right) = p$ et $V\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{n}$.

3. Soit $\varepsilon > 0$. Démontrer que

$$\mathrm{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant \frac{1}{4n\varepsilon^2}$$

✓ D'après l'inégalité de Bienaymé-Tchebychev,

$$\mathrm{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2}$$

et on conclut en remarquant que $p(1-p) \leqslant \frac{1}{4}$.

4. Comment choisir la taille de l'échantillon de sorte que le résultat du sondage soit proche de p à $\varepsilon = 0.05$ près avec une probabilité supérieure à 95% ?

✓ Il faut choisir n tel que $\frac{1}{4}n\varepsilon^2 \leqslant 0.05$, on peut prendre $n = 2000$.