feuille 2 exo $\overline{XIV}$

on note N_A le nombre aléatoire d'électeurs qui choisiront le candidat A

n le nombre d'électeurs ds l'échantillon.

F_A la fréquence aléatoire ~~des~~ choix pour A $\quad F_A = \frac{N_A}{n}$

X_i la v.a qui vaut 1 si l'électeur "i" choisit A et qui vaut 0 sinon.

$Prob(X_i = 1) = Prob(\text{le } i^e \text{ électeur choisit } A) = \frac{6695000}{13000000} = 0{,}515 = p$

$Prob(X_i = 0) = 1 - p = 0{,}485 = q$

$X_i \sim B(0{,}515)$ (loi de Bernoulli de paramètre 0,515)

$E(X_i) = 0{,}515$ et $Var(X_i) = pq = 0{,}515 \times 0{,}485 = 0{,}249775$

$\sigma_A = 0{,}499775$

Remarquons que $N_A = \sum_{i=1}^{n} X_i$

N_A est la somme de n variables de Bernoullis indépendants de même paramètre 0,515 donc N_A suit une loi binomiale de paramètre n et p ; $N_A \sim B(n, 0{,}515)$

Remarquons que $F_A = \frac{N_A}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}_n$

On nous demande de chercher la probabilité pour que l'institut se trompe c'est à dire ~~que~~ la probabilité de déclarer que B gagne (alors que c'est A qui gagnera) ; c'est à dire on cherche la probabilité pour que $N_A < (n - N_A)$ alors que $N_A \sim B(n, 0{,}515)$

$Prob(N_A < n - N_A) = Prob(2N_A < n) = Prob(N_A < \frac{n}{2})$

$= Prob(B(n ; 0{,}515) < \frac{n}{2})$

Si on fait l'hypothèse a priori que $n > 30$, $np > 15$, $nq > 15$ et comme $p = 0{,}515$ pas trop proche de 0 ni de 1 on peut approximer $B(n ; 0{,}515)$ par $\mathcal{N}(n \times 0{,}515 \,;\, n \times 0{,}515 \times 0{,}485)$. A posteriori $n = 100$ ou $n = 1600$; les conditions d'approximation sont valides.

Ainsi : $Prob(N_A < n - N_A) = Prob(N(0{,}515n \,;\, 0{,}249775\, n) < \frac{n}{2})$

$= Prob(N(0,1) < \frac{\frac{n}{2} - 0{,}515n}{0{,}499775\sqrt{n}})$ après centrage et réduction

$= Prob(N(0,1) < \frac{-0{,}015\sqrt{n}}{0{,}499775})$ d'où pour $n = 100$; $Prob(N_A < n - N_A) = 38{,}2\%$

pour $n = 1600$, $Prob(N_A < n - N_A) = 11{,}5\%$

Feuille 2 TD XIV suite et fin

Conclusion : On constate que plus l'échantillon est grand en taille plus le risque d'erreur diminue. Ce risque passe de 38% à 11,5% lorsqu'on passe d'un échantillon de taille 100 à 1600.

Ce phénomène est à associer avec la loi faible des grands nombres. Cette loi permet d'affirmer que sous certaines conditions ~~sur~~ une suite de v.a $X_1, X_2, \dots X_n, \dots$, la variable aléatoire $\overline{X}_n = \frac{N_A}{n} = F_A$ converge en probabilité vers 0 quand $n \longrightarrow \infty$, c'est à dire, pour un intervalle donné autour de la ~~moyenne~~ espérance $E(X_i)$, lorsque l'effectif n augmente, la probabilité pour que $\overline{X}_n$ tombe à l'extérieur de cet intervalle devient de plus en plus faible.

Autrement dit, le risque d'erreur de déclarer que F_A ou $\frac{N_A}{n}$ soit loin de la ~~moyenne~~ espérance $E(X_i)$ diminue quand la taille de l'échantillon augmente. En l'occurrence, dans notre exemple, si $E(X_i) = 0,515$, (~~ce~~ qui signifie que A gagne) déclarer que B gagnerait correspond à $F_A < 0,5$ (soit F_A se placerait à l'extérieur de $]0,5 ; 0,530[$) aurait une probabilité de plus en plus petite lorsque la taille augmente. Le risque d'erreur diminue.