Présidentielle 2022 : Probabilités d'accès au 2nd tour

Note méthodologique

Pour chacun de ces sondages, nous modélisons le score de chacun des n candidats par une variable aléatoire X_i avec une espérance $\bar{X}_i$ égale au score effectif du candidat dans ce sondage, avec $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \bar{X}_i = 1$.

Nous avons donc $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix} + \begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_n - \bar{X}_n \end{pmatrix}$ avec le premier vecteur $\begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$ entièrement connu et le deuxième vecteur $\begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_n - \bar{X}_n \end{pmatrix}$ devant satisfaire la condition $\sum_{i=1}^{n}(X_i - \bar{X}_i) = 0$.

Ainsi, en connaissant le score simulé de $n-1$ candidats, on peut en déduire le score du $n^{\text{ème}}$ candidat : ce vecteur a $n-1$ degrés de liberté.

Mathématiquement, cela traduit que la projection orthogonale du vecteur aléatoire $\begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_n - \bar{X}_n \end{pmatrix}$ sur le sous-espace de dimension 1 défini par le vecteur $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ est nulle, et que le vecteur aléatoire $\begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_n - \bar{X}_n \end{pmatrix}$ appartient au sous-espace complémentaire de dimension $n-1$.

Notons $\vec{X}$ le vecteur aléatoire $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $\vec{\bar{X}}$ le vecteur constant $\begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$, $\overrightarrow{X - \bar{X}}$ le vecteur aléatoire $\begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_n - \bar{X}_n \end{pmatrix}$, $\vec{1}$ le vecteur constant $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ et $\vec{N}$ le vecteur aléatoire $\begin{pmatrix} N_1 \\ \vdots \\ N_n \end{pmatrix}$ avec $N_i \sim N(0, \sigma_i)$ (variable aléatoire suivant une loi normale d'espérance nulle et d'écart type σ_i, σ_i étant l'écart-type correspondant à la marge d'erreur du sondage pour le candidat i).

Le vecteur $\left(\vec{N} - \frac{\vec{1}.\vec{N}}{n^2}\vec{1}\right)$ est la projection du vecteur $\vec{N}$ sur le sous-espace de dimension $n-1$ complémentaire au sous-espace de dimension 1 défini par le vecteur $\vec{1}$. Nous avons donc $\overrightarrow{X - \bar{X}} \sim \left(\vec{N} - \frac{\vec{1}.\vec{N}}{n^2}\vec{1}\right)$ et $\vec{X} \sim \left(\vec{N} - \frac{\vec{1}.\vec{N}}{n^2}\vec{1} + \vec{\bar{X}}\right)$.

Nous pouvons donc simuler les scores des candidats en simulant des réalisations de la variable aléatoire $\left(\vec{N} - \frac{\vec{1}.\vec{N}}{n^2}\vec{1} + \vec{\bar{X}}\right)$, avec $N_i \sim N(0, \sigma_i)$.