

**PROJET SEMESTRIEL : DÉCOUPAGE
ÉLECTORAL**

ISRAA BEN SASSI, CAMILLE GOUJET ET VICTOR
LESUR

À L'ATTENTION DE ARNAUD KNIPPEL

FIGURE 0.0.1. Illustration du découpage électoral en France

Contents

Part 1. PARTIE THÉORIQUE	5
Chapter 1. Introduction	6
1.1. Histoire	6
1.2. Quelques définitions importantes	6
Chapter 2. Modéliser le découpage	8
2.1. Application sur l'exemple donné	8
2.2. Modélisation du découpage	9
Chapter 3. La classification des données	13
3.1. Intérêt	13
3.2. Classification non supervisée	13
Part 2. PARTIE PRATIQUE	14
Chapter 4. Utilisation d'un solveur	15
Chapter 5. Algorithme de classification non supervisée	16

ABSTRACT. Le découpage électoral est le mécanisme par lequel le territoire national est subdivisé en circonscriptions électorales, destinées à permettre et favoriser l'expression des citoyens par le vote.

Ce découpage est l'objet récurrent de polémiques, sinon d'actions en justice. Car il est orchestré, par exemple aux Etats Unis, au niveau de chaque État par les gouverneurs et ceux-ci ont la fâcheuse tendance à faire en sorte de favoriser leur parti. Par exemple, en façonnant des circonscriptions qui fragmentent l'électorat de leur adversaire, pour limiter son nombre d'élus. Ou, au contraire, en rassemblant dans une seule et même circonscription un maximum de votes adverses : ainsi, même avec 80% des voix, cela ne fera jamais qu'un seul élu dans la circonscription. C'est ce que l'on appelle gâcher des voix. Il arrive ainsi qu'à l'échelle d'un pays, un parti a moins d'élus qu'un autre alors qu'il a eu plus de voix. Un exemple? Une certaine Hillary Clinton a perdu la présidentielle en 2016 avec plus de 2 millions de votes d'avance sur son adversaire.

Ce projet consiste à étudier une ou plusieurs méthodes de découpage électoral, et à essayer de comprendre et démontrer dans quelle mesure on peut influencer le résultat d'élections par ce découpage.

Dans un premier temps, nous nous intéresserons à l'histoire de ce mécanisme et redéfinirons quelques notions de base pour que la compréhension du lecteur pour le reste du projet soit optimale.

Nous étudierons ensuite à partir de situations simples et faciles à modéliser de petites bases de données. À partir de celles-ci, nous modéliserons des programmes linéaires en nombres entiers que nous tenterons de résoudre pour en mesurer l'écart des résultats possibles et donc la manipulation partisane que l'on peut en faire pour un camp ou un autre..

Nous généraliserons nos modèles à des bases de données de plus en plus conséquentes et complexes et en étudierons les aspects.

Enfin, nous tenterons de remédier au problème d'équité qui pèse sur les partis à chaque nouvelle élection en élaborant une modélisation la plus équitable et représentative possible.

Afin de mener à bien ce projet, nous nous munirons d'un solveur tel que CPLEX pour nous assister dans la résolution de problèmes d'optimisation linéaire.

Part 1

PARTIE THÉORIQUE

CHAPTER 1

Introduction

1.1. Histoire

Le découpage électoral partisan des circonscriptions, aussi connu sous le nom de « *gerrymandering* », est une pratique qui n'est pas récente et qui est sûrement concomitante d'élections de représentants dans ces circonscriptions. Néanmoins, c'est avec l'apparition des démocraties représentatives modernes, d'abord les États-Unis puis la France, que la question est mise sur le devant de la scène. Ainsi, le terme *gerrymandering* nous vient des USA et fait référence à un découpage effectué par le gouverneur Elbridge Gerry. Ce découpage était si bizarre qu'il rappelait la forme d'une salamandre, d'où le mot-valise *gerrymandering* (composé de Gerry et de salamander du français salamandre).

De nos jours avec des outils de recensement d'une plus grande précision et des outils informatiques puissants, le découpage partisan fait toujours débat. En France, par exemple, il y a eu trois découpages électoraux sous la cinquième République (soit en 1958, 1986 et 2010). Dans ces trois cas, le gouvernement n'a nécessité uniquement d'une majorité absolue au parlement et non d'une majorité qualifiée (celle-ci est nécessaire pour un changement de constitution par exemple)¹. Ainsi, il est plus facile au gouvernement de procéder au *gerrymandering*. Néanmoins, le gouvernement ne peut faire ce qu'il veut. En France, tout découpage électoral doit être sur une base essentiellement démographique, comme l'indique le conseil constitutionnel². De fait, comme nous le verrons plus tard, le nombre d'habitants d'une circonscription doit être compris entre 80 et 120% de la moyenne des circonscriptions de la même région.

1.2. Quelques définitions importantes

DÉFINITION 1. Découpage électoral

Le découpage électoral est le mécanisme par lequel le territoire national est subdivisé en *circonscriptions électorales*, destinées à permettre et favoriser l'expression des citoyens par le vote.

REMARQUE. En France, le mécanisme du découpage électoral est utilisé pour les élections législatives, les élections sénatoriales et les élections au Conseil départemental.

DÉFINITION 2. Circonscription électorale

La circonscription électorale est une division du territoire effectuée dans le cadre d'une élection.

Chaque citoyen est rattaché à une circonscription et à une seule dans le cadre d'un vote. La circonscription peut être utilisée dans le cadre des scrutins majoritaires ou dans le cadre des scrutins de liste (scrutins proportionnels).

EXEMPLE. 10 000 électeurs, 10 circonscriptions de 1 000 électeurs, deux partis A et B, le premier recueillant 4 000 suffrages et le second 6 000.

- 6 circonscriptions gagnées par le parti A avec 53,3% des voix soit 3200 voix pour A et 2800 voix pour B
- 4 circonscriptions gagnées par le parti B avec 80% des voix, soit 800 voix pour A et 3200 voix pour B.

Le parti B perd les élections bien qu'il soit largement majoritaire en nombre de voix.

CHAPTER 2

Modéliser le découpage

2.1. Application sur l'exemple donné

Grâce à l'exemple ci-dessus, on remarque que le découpage électoral peut grandement influencer le résultat d'une élection. Bien modéliser ce découpage est alors essentiel; que notre but soit de favoriser un camp ou d'essayer de créer des conditions les plus équitables possible. Pour commencer, nous allons vous proposer la modélisation suivante :

Chaque circonscription se décompose en "brique" élémentaire que sont les cantons. Ces cantons seront dans notre modèle assimilés à des nœuds d'un graphe. Chaque canton est relié par des arêtes aux autres cantons limitrophes. Pour simplifier, nous allons considérer seulement deux partis A et B. Ainsi, on attribua un poids à chaque nœud correspondant à un réel entre 0 et 1 qui traduit le pourcentage de votants qui soutiennent le parti A. Donc si le poids est de 0, tous les votants soutiennent le parti B et si le poids est de 1, tous les votants soutiennent le parti A.

Créer les circonscriptions à partir du graphe revient à résoudre un problème de classification non supervisée. En effet, chaque circonscription ayant un seul élu, on peut influencer le nombre d'élus obtenu par un parti en découpant différemment les circonscriptions.

Pour rajouter un minimum de réalisme, on peut exiger que les circonscriptions soit d'un seul tenant, c'est-à-dire rajouter une **contrainte de connexité entre les nœuds d'une même classe.**

Si l'on reprend l'exemple précédent de **10 000 électeurs** à répartir sur **10 circonscriptions de 1 000 électeurs**, on peut imaginer que les électeurs se répartissent sur des **cantons de 200 électeurs** où 16 cantons soutiennent le parti B (poids 0), 28 soutiennent le parti A (poids 1) et 6 sont des cantons de poids 0,33.

Alors on peut répartir ces cantons de façon à obtenir :

- 4 circonscriptions avec un canton B et 4 cantons A
- 6 circonscriptions avec 2 cantons A, 2 cantons B et un canton de poids 0,33.

Avec cette répartition, le parti B remporte une majorité de circonscriptions sans pour autant remporter le plus grand nombre de voix. Cette répartition est représentée par le graphe ci-dessous :

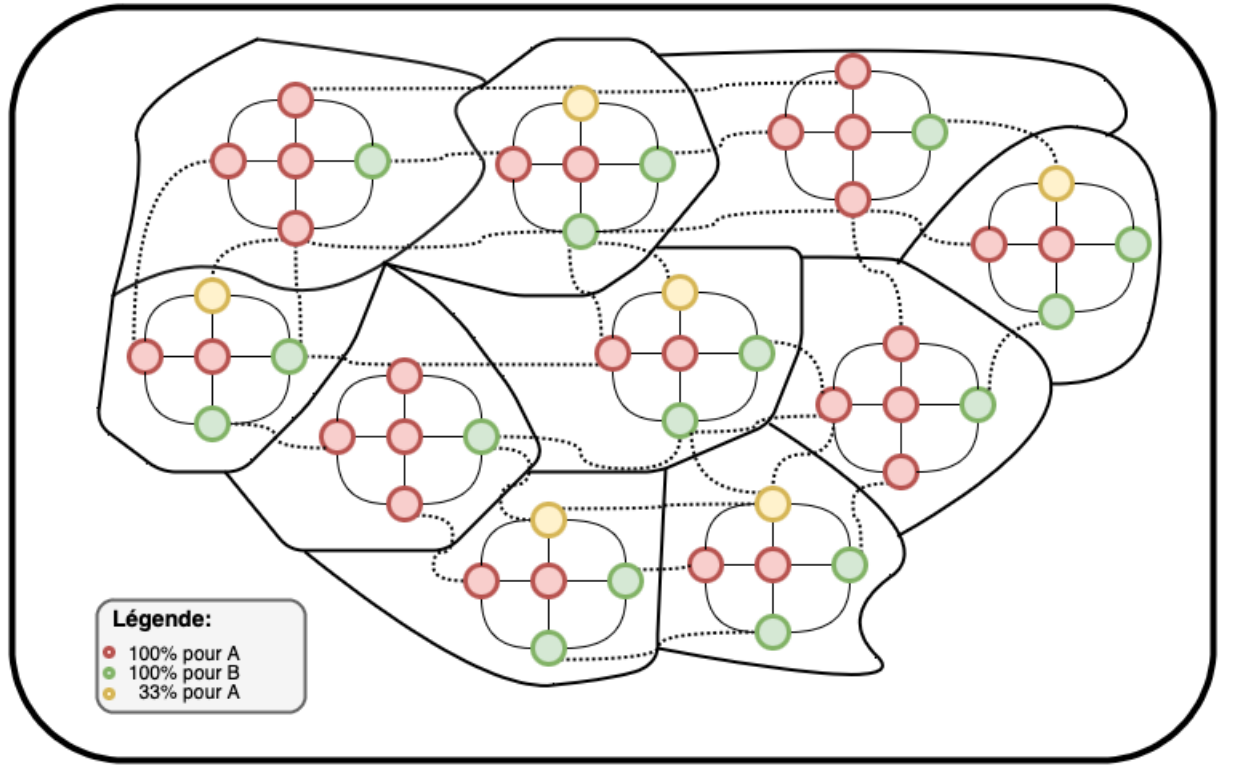


FIGURE 2.1.1. Graphe modélisant la situation

2.2. Modélisation du découpage

2.2.1. Problème d'optimisation.

DÉFINITION 3. Un problème d'optimisation combinatoire est défini par :

- un énoncé
- une fonction économique
- un sens.

2.2.1.1. Énoncé:

Dans notre cas, l'énoncé, c'est-à-dire les données, correspond au graphe $G = (X, U)$ décrit ci-dessus. Soit $X = \{x_1, x_2, \dots, x_n\}$ l'ensemble des nœuds de notre graphe, tel que chaque x_i correspond à un canton. On lui associe les variables nbV_i, pV_i qui sont respectivement le **nombre de votants du canton** et le **pourcentage de votants qui soutiennent le parti A**.

. Notre objectif est de définir un nombre k de circonscriptions tel que $\forall j \in \{1, \dots, k\}$ K_j est un ensemble connexe de X , où un élément de X n'appartient qu'à un seul K_j .

Plaçons nous dans un premier temps dans la situation où notre objectif est de **faire gagner le parti A** :

Nous pouvons commencer par définir la fonction p qui calcule le **poids du parti A dans une circonscription donné, soit la probabilité de votants pour le parti A dans cette circonscription.**

Le résultat $p(K_j)$ nous donne en fait le rapport du nombre de personnes votant pour le parti A dans la circonscription K_j sur le nombre total de votants dans cette même circonscription :

$$\begin{aligned} p : X &\rightarrow \mathbb{R}^+ \\ K_j &\mapsto p(K_j) \\ p(K_j) &= \frac{\sum_{i=1, x_i \in K_j}^n nbV_i \times pV_i}{\sum_{i=1, x_i \in K_j}^n nbV_i} \end{aligned}$$

2.2.1.2. La fonction économique.

Ceci nous permet de définir fonction économique f suivante :

$$\begin{aligned} f : X &\rightarrow \mathbb{R}^+ \\ K_j &\mapsto f(K_j) \\ f(K_j) &= \begin{cases} 1 & \text{si } p(K_j) \geq 0,5 \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Ainsi la fonction f nous dit si c'est le parti A remporte la circonscription K_j ($f(K_j) = 1$) ou si c'est le parti B qui remporte la circonscription ($f(K_j) = 0$).

2.2.1.3. *Sens.* Si notre objectif est de faire gagner le plus de circonscriptions au parti A, on peut définir le problème suivant :

$$\max \sum_{j=1}^k f(K_j)$$

Inversement, si notre objectif est de faire gagner le plus de circonscriptions au parti B, il suffit de poser :

$$\min \sum_{j=1}^k f(K_j)$$

Remarquons que dans notre formulation, nous avons posé une contrainte de façon à ce que les circonscriptions soient d'un seul tenant mais nous n'avons pas imposé de contrainte de taille sur les circonscriptions.

Pour cela, nous pouvons nous inspirer des contraintes qui existent sur la taille des circonscriptions en France.

La règle est que le nombre de votants d'une circonscription doit être

compris entre 80% et 120% du nombre moyen de votants dans une circonscription.

On peut ainsi définir la contrainte suivante pour un K_j donné :

$$\frac{0,8}{k} \sum_{i=1}^n nbV_i \leq \sum_{i=1, x_i \in K_j}^n nbV_i \leq \frac{1,20}{k} \sum_{i=1}^n nbV_i$$

Notons aussi que si notre objectif n'est pas de favoriser un camp ou un autre, alors l'utilisation de la fonction f n'a pas d'intérêt.

Il faudrait définir une nouvelle fonction économique tout en gardant la contrainte sur la taille que l'on vient d'énoncer.

2.2.2. Le choix de variables

. Afin de pouvoir formuler le problème sous forme d'un problème mathématique nous avons besoin de définir les variables que nous allons utiliser. Définissons simplement les variables binaires x_{ij} tel que pour le $i^{\text{ème}}$ canton et la $j^{\text{ème}}$ circonscription on ai :

$$x_{ij} = \begin{cases} 1 & \text{si } x_i \in K_j \\ 0 & \text{sinon} \end{cases}$$

On peut remarquer que cette variable traduit l'appartenance d'un canton à une circonscription. Ce qui implique de devoir énoncer ces deux contraintes suivantes :

$$\begin{aligned} \forall i \sum_{j=1}^k x_{ij} &= 1 \\ \forall j \in \{1, \dots, k\} \sum_{i=1}^n x_{ij} &\geq 1 \end{aligned}$$

La première imposant à chaque canton d'appartenir à une et une seule circonscription et la seconde obligeant chaque circonscription à contenir au-moins un canton. Cet dernière contrainte peut par ailleurs être modifiée pour forcer les circonscriptions à chacune contenir, un nombre de canton que l'on pourra définir.

En explicitant les variables de la sorte, on peut alors redéfinir la fonction $p(K_j)$ précédente

$$p(K_j) = \frac{\sum_{i=1}^n x_{ij} \times nbV_i \times pV_i}{\sum_{i=1}^n x_{ij} \times nbV_i}$$

On peut donc formuler notre problème mathématique

$$\begin{aligned} \max(z) = & \sum_{j=1}^k \left\lfloor 2 \left(\frac{\sum_{i=1}^n x_{ij} \times nbV_i \times pV_i}{\sum_{i=1}^n x_{ij} \times nbV_i} \right) \right\rfloor \\ & \begin{cases} \sum_{j=1}^k x_{ij} = 1 & \forall i \in \{1, \dots, n\} \\ \sum_{i=1}^n x_{ij} \geq 1 & \forall j \in \{1, \dots, k\} \\ \sum_{i=1}^n x_{ij} \times nbV_i \leq \frac{1,20}{k} \sum_{i=1}^n nbV_i & \forall j \in \{1, \dots, k\} \\ \sum_{i=1}^n x_{ij} \times nbV_i \geq \frac{0,8}{k} \sum_{i=1}^n nbV_i & \forall j \in \{1, \dots, k\} \\ x_{ij} \in \{0, 1\} & \forall (i, j) \in \{1, \dots, n\} \times \{1, \dots, k\} \end{cases} \end{aligned}$$

La dernière contrainte qui nous manque est celle de la connexité. C'est-à-dire que pour tout couple de sommets appartenant à une même classe, il existe un chemin pour aller de l'un à l'autre ne contenant que des sommets de leur classe. Pour expliciter cette notion, posons $G_j = (K_j, U_j)$ où K_j est l'ensemble des sommets d'une circonscription et U_j est l'ensemble des arcs reliant les sommets de K_j , alors nous voulons que $\forall j \in \{1, \dots, k\}$ et $\forall (x_i, x_{i'}) \in K_j^2$ $x_i \longleftrightarrow x_{i'}$ dans G_j .

CHAPTER 3

La classification des données

3.1. Intérêt

3.2. Classification non supervisée

3.2.1. Principe.

3.2.2. Application.

Part 2

PARTIE PRATIQUE

CHAPTER 4

Utilisation d'un solveur

CHAPTER 5

Algorithme de classification non supervisée