

Modélisation de la distribution des erreurs de comptage pour les élections communales en Wallonie en 2018.

3 septembre 2019

1 Contexte

Pour les élections communales belges de 2018, une grande majorité des bureaux de vote en Wallonie organise le scrutin en vote papier. Les bulletins sont déposés dans des urnes qui sont acheminées vers des bureaux de dépouillement où les bulletins sont alors comptés par des citoyens assignés à cette tâche. Il est tout naturel de considérer qu'à chaque fois qu'un bulletin de vote individuel est compté, il peut survenir une erreur de comptage de celui-ci (nous noterons la probabilité qu'une telle erreur survienne ϵ). Pour notre étude, nous définissons cette erreur comme ceci : un vote compté pour un parti alors qu'il aurait dû être comptés pour un autre. Pour la suite nous considérons comme un parti à part entière, l'ensemble des votes blancs et nuls qui sont agrégés en Wallonie.

La probabilité ϵ va dépendre de plusieurs facteurs dont par exemple, la qualité de l'environnement de travail du bureau de dépouillement, les compétences de la personne qui compte, ou encore si cette personne est assistée dans le décompte par d'autres personnes voire par des machines spécialisées. Nous n'avons aucune idée de ce que ϵ peut valoir dans chaque cas de figure et nous ne connaissons pas non plus de valeur moyenne pour ϵ .

On peut s'interroger sur cette probabilité ϵ de commettre une erreur, mais faute de données à ce sujet, on se contentera d'étudier *le taux d'erreur global* du décompte d'un scrutin que nous noterons α . Ce taux d'erreur α indique la différence entre le résultat issu du dépouillement et le résultat correct. Il est une conséquence directe des erreurs individuelles sur les bulletins et donc de ϵ . Pour pouvoir calculer α , il faudrait pouvoir disposer du résultat correct, ce que nous n'avons pas. Néanmoins, nous pouvons estimer le taux d'erreur α à partir des recomptages effectués après le premier décompte¹.

1. Les taux d'erreur que nous constatons sont donc eux-même entachés d'erreur. Néanmoins, on peut espérer que cette deuxième erreur soit plus petite car les bulletins recomptés ont déjà été triés par le premier décompte. Il s'agit donc ici d'un travail de détection d'intrus dans une pile de bulletins plutôt que de triage avec décompte, ce qui semble moins propice aux erreurs.

Pour le taux d'erreur α , nous avons heureusement un peu plus de données à disposition même si cela reste maigre car il existe peu de cas bien documentés de recomptage des votes papiers en Wallonie et en Belgique. Toutefois, les sources² dont nous disposons semblent indiquer une valeur du taux d'erreur comprise entre 0,5% et 2% dans un scénario générique (loin du contexte belge). Les quelques cas belges recensés font part d'une erreur entre $\sim 0,1\%$ et $\sim 1,5\%$.

C'est dans ce contexte que nous avons voulu quantifier ce que ces erreurs de comptage auraient comme conséquences sur les scores particuliers des partis. Autrement dit, comment un taux d'erreur global se répercute au niveau du score des partis et des votes blancs et nuls.

2 Description de l'expérience

Pour notre expérimentation, nous avons simulé des erreurs survenant lors du décompte pour un nombre fixé de bulletins choisis au hasard. Pour ce faire, nous avons procédé comme ceci :

1. on sélectionne un bulletin de vote B_1 sur l'ensemble des bulletins soumis de manière uniformément aléatoire. Si ce bulletin a déjà été modifié par l'étape 3, on recommence en tirant au sort un autre bulletin jusqu'à trouver un bulletin non encore modifié.
2. on sélectionne un autre bulletin de vote B_2 sur l'ensemble des bulletins soumis de manière uniformément aléatoire.
3. si les votes contenu dans B_1 et B_2 sont identiques, on retourne à l'étape 2. Si ils sont différents, on transforme le vote du bulletin B_1 en un vote similaire au vote du bulletin B_2 .

On recommence ce procédé un nombre de fois équivalent à $\alpha\%$ des bulletins soumis. On obtient alors un nouveau décompte potentiellement différent du décompte initial. On peut alors noter les différences obtenues entre les deux décomptes, c'est-à-dire les différences de scores pour chaque parti entre les deux décomptes.

Pour être plus parlant, voici l'exemple d'une simulation d'erreur de décompte obtenue pour le scrutin de la commune de Doische en modifiant $\alpha = 1\%$ des bulletins de vote. Sur un total de 2200 bulletins soumis, cela représente 22 bulletins modifiés. On peut visualiser les résultats de cette simulation sous la forme d'un tableau (Fig. 1) où les entrées indiquent le nombre de votes qui transitent d'un parti à l'autre (pour rappel, blancs et nuls sont considérés comme un parti). Par exemple, 4ème ligne, 2ème colonne, la liste ENSEMBLE va "donner" 6 voix à la liste MR-IC. La ligne *total* reprend les transferts totaux pour chaque parti soit la somme de chaque colonne. Les deux dernières lignes reprennent les scores des partis avant et après les transferts de votes.

On peut à présent répéter cette expérience un grand nombre de fois pour obtenir les valeurs moyennes du nombre de voix qui vont rentrer ou sortir d'un

2. Voir les références dans le post du blog decryptage.be.

	→ MR-IC	→ CD@	→ ENSEMBLE	→ Blancs et nuls
MR-IC →	-8	+2	+6	0
CD@ →	0	-1	+1	0
ENSEMBLE →	+6	+3	-10	+1
Blancs et nuls →	+1	+1	+1	-3
<i>total</i>	-1	+5	-2	-2
ancien score	1206	260	637	97
nouveau score	1205	265	635	95

FIGURE 1 – Résultat d’une simulation pour la commune de Doische en considérant 1% d’erreur sur l’ensemble des 2200 bulletins soumis.

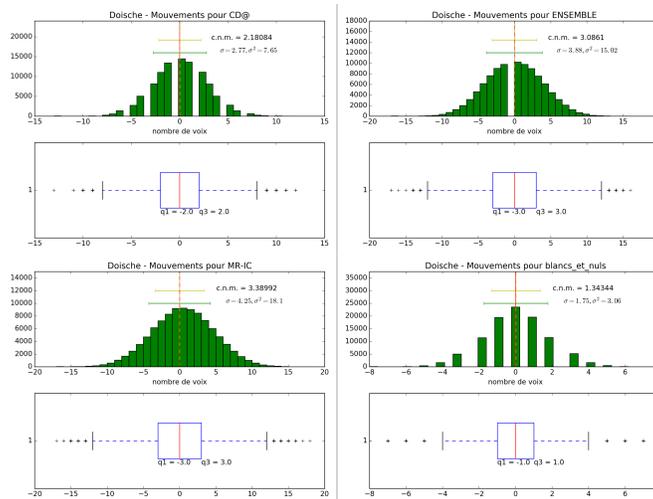


FIGURE 2 – Résultats des simulations pour la commune de Doische. Mouvements pour les quatre partis.

parti. On pourra alors mesurer les tendances générales ainsi que leur dispersion autour de la moyenne grâce à la variance et l’écart-type.

3 Résultats des simulations

Pour chaque commune, nous avons effectué $n = 100\,000$ simulations comme celle décrite à la section précédente. Pour chaque parti de chaque commune nous pouvons alors calculer les changements nets moyens (c.n.m.) : $\frac{1}{n} \sum_{i=1}^n |\Delta_i|$ où Δ_i est la différence entre le nouveau et l’ancien score. On peut également calculer l’écart-type σ et la variance σ^2 ainsi que les valeurs médianes, 1er et 3ème quartiles ($q1$ et $q3$). Voici ce que l’on obtient par exemple pour la commune de Doische (Fig. 2).

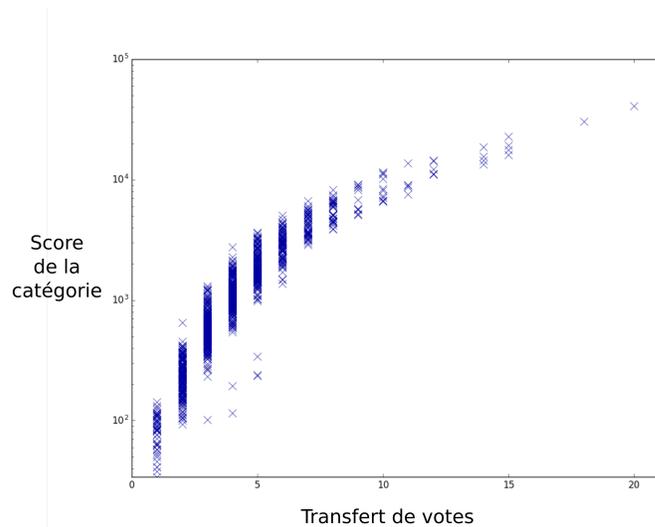


FIGURE 3 – Scores de toutes les partis de toutes les communes wallonnes (en Oy) sur le changement net moyen arrondi à l'entier (Ox). Par exemple un parti réalisant un score de ± 10.000 votes subit un transfert de $\pm 9, \pm 10, \pm 11$ ou ± 12 votes en moyenne.

Ces graphiques et ces valeurs nous permettent de visualiser clairement le type de répartition des erreurs ainsi que leur fréquence. On peut voir par exemple pour la commune de Doische, pour le parti MR-IC, une distribution centrée autour de la moyenne et de la médiane (ici 0) avec un écart-type de 4,25 votes. Cela nous indique que les erreurs globales de 1% correspondent en général pour ce parti à un transfert de plus ou moins 0,1,2,3 ou 4 votes. Les résultats de toutes ces simulations sont disponibles ici ??.

Les communes wallonnes possèdent en général des profils similaires. On observe, par exemple, de grosses communes (en termes d'électeurs) comptant de nombreuses listes dont quelques unes réalisant de gros scores et beaucoup réalisant de petits scores ou au contraire, de petites communes comptant peu de listes avec des scores similaires. Les profils que l'on n'observe peu sont de grosses communes avec peu de listes ou de petites communes avec beaucoup de listes. En partant de ce constat, on peut mettre en lumière une tendance entre le score d'un parti et le nombre de votes qu'il va en moyenne perdre ou recevoir selon le taux d'erreur global. Dans la Figure 3, on a mis en rapport les scores de toutes les partis de toutes les communes wallonnes sur le changement net moyen arrondi pour un taux d'erreur global $\alpha = 1\%$. On observe cette tendance qui est liée à la similarité des profils entre communes.

4 Analyse statistique

Dans cette section, nous dégagerons les lois statistiques qui régissent l'expérience et dont les résultats concordent avec les résultats obtenus par les simulations (voir tous les résultats ici ??).

NOTATIONS :

S_i : score du parti i , c'est-à-dire le total des votes attribués au parti i avec potentiellement des erreurs d'attribution.

N : nombre de partis.

$S_T = \sum_{l=1}^N S_l$: score total des partis.

X_{ij} : variable aléatoire décrivant le nombre de votes pour le parti j incorrectement attribués au parti i (donc dans le score de S_i et non S_j). On peut lire X_{ij} comme «le nombre de votes comptés pour le parti i qui devraient être comptés pour le parti j ».

Δ_i : variation du score du parti i après recomptage des votes et éliminations des erreurs d'attribution.

ϵ : probabilité pour un vote qu'une erreur de comptage se produise.

K : variable aléatoire qui représente le nombre total de vote mal comptés.

$K \sim$ loi binomiale (S_T, ϵ) avec pour fonction de densité

$$Pr[K = k] = C_{S_T}^k \epsilon^k (1 - \epsilon)^{S_T - k}.$$

E : espérance.

V : variance.

COV : covariance.

On pose la contrainte suivante : $\sum_{i,j=1;i \neq j}^N X_{ij} = k$ ce qui nous amène à nous intéresser à la fonction de densité

$$Pr[X_{ij} = x_{ij}, i, j = 1, \dots, N; i \neq j | \sum X_{ij} = k].$$

Cette fonction suit une loi multinomiale et pour un k fixé, on obtient

$$\begin{aligned} & Pr[X_{ij} = x_{ij}, i, j = 1, \dots, N; i \neq j | \sum x_{ij} = k] \cdot Pr[K = k] \\ &= \frac{k!}{\prod_{i,j=1;i \neq j}^N x_{ij}} \cdot \prod_{i,j=1;i \neq j}^N \left(\frac{S_i}{S_T} \cdot \frac{S_j}{S_T - S_i} \right)^{x_{ij}} \cdot Pr[K = k]. \end{aligned}$$

Pour cette loi multinomiale on peut estimer les valeurs de dispersion suivantes

pour les variables X_{ij} :

$$\begin{aligned} E(X_{ij}) &= k \cdot \frac{S_i}{S_T} \cdot \frac{S_j}{S_T - S_i}, \\ V(X_{ij}) &= k \cdot \frac{S_i}{S_T} \cdot \frac{S_j}{S_T - S_i} \left(1 - \frac{S_i}{S_T} \cdot \frac{S_j}{S_T - S_i}\right), \\ COV(X_{ij}, X_{lm}) &= -k \cdot \left(\frac{S_i}{S_T} \cdot \frac{S_j}{S_T - S_i}\right) \cdot \left(\frac{S_l}{S_T} \cdot \frac{S_m}{S_T - S_l}\right). \end{aligned}$$

In fine, ce qui nous intéresse, c'est la variation du score d'un parti que l'on note Δ_i :

$$\begin{aligned} \Delta_i &= \sum_{j=1, j \neq i}^N X_{ji} - \sum_{j=1, j \neq i}^N X_{ij} \\ &= \sum_{j=1, j \neq i}^N (X_{ji} - X_{ij}). \end{aligned}$$

Pour un k fixé on aura alors,

$$\begin{aligned} E(\Delta_i | K = k) &= E\left(\sum_{j=1, j \neq i}^N (X_{ji} - X_{ij})\right) \\ &= \sum_{j=1, j \neq i}^N (E(X_{ji}) - E(X_{ij})), \\ V(\Delta_i | K = k) &= V\left(\sum_{j=1, j \neq i}^N (X_{ji} - X_{ij})\right) \\ &= \sum_{j=1, j \neq i}^N V(X_{ji} - X_{ij}) + \sum_{j, l=1; j, l \neq i}^N COV(X_{ji} - X_{ij}, X_{li} - X_{il}) \\ &= \sum_{j=1, j \neq i}^N (V(X_{ji}) + V(X_{ij}) - 2COV(X_{ji}, X_{ij})) \\ &\quad + \sum_{j, l=1; j, l \neq i}^N (COV(X_{ji}, X_{li}) - COV(X_{ji}, X_{il}) - COV(X_{ij}, X_{li}) \\ &\quad + COV(X_{ij}, X_{il})). \end{aligned}$$

Dans le cas général où k non fixé – c'est-à-dire que k prend des valeurs entre 0 et S_T – on aura

$$\begin{aligned} E(\Delta_i) &= \sum_k E(\Delta_i | K = k) \cdot Pr[K = k], \\ V(\Delta_i) &= \sum_k (V(\Delta_i | K = k) + E(\Delta_i | K = k)^2) \cdot Pr[K = k] - E(\Delta_i)^2. \end{aligned}$$

On obtient la deuxième égalité à partir de

$$V(\Delta_i) = E(\Delta_i^2) - E(\Delta_i)^2 \quad (1)$$

et en observant que

$$\begin{aligned} \sum_k V(\Delta_i|K = k) \cdot Pr[K = k] &= \sum_k E((\Delta_i - E(\Delta_i|K = k))^2|K = k) \cdot Pr[K = k] \\ &= \sum_k E(\Delta_i^2|K = k) \cdot Pr[K = k] \\ &\quad - 2E(\Delta_i \cdot E(\Delta_i|K = k)|K = k) \cdot Pr[K = k] \\ &\quad + E(\Delta_i|K = k)^2 \cdot Pr[K = k] \\ &= \sum_k E(\Delta_i^2|K = k) \cdot Pr[K = k] \\ &\quad - E(\Delta_i|K = k)^2 \cdot Pr[K = k] \\ &= E(\Delta_i^2) - \sum_k E(\Delta_i|K = k) \cdot Pr(K = k). \end{aligned}$$

En effet, cette dernière égalité nous permet de substituer $E(\Delta_i^2)$ dans (1) et d'obtenir ainsi la formule pour $V(\Delta_i)$.

5 Conclusion

Dans cet article, nous avons étudié la manière dont les erreurs de comptage lors des élections papier en Wallonie se répercutent sur les scores individuels des partis. Nous avons constaté qu'une erreur de comptage globale (par exemple de l'ordre de 1%) se traduisait par une variation du score au sein de chaque liste dont on peut calculer la distribution avec une bonne précision. Nous avons simulé cette distribution en faisant apparaître aléatoirement des erreurs lors du décompte et nous avons établi les lois statistiques qui sous-tendent à cette distribution.

Malheureusement, nous avons peu de données pour conforter nos hypothèses initiales. Nous ne savons quasiment rien sur la probabilité avec laquelle une erreur de comptage se produit en moyenne en Wallonie. Et nous savons peu de choses sur le taux d'erreur global d'un décompte. Les quelques données issues des recomptages qui ont eu lieu lors d'élections précédentes nous fournissent une indication sur l'ordre de grandeur de ce taux d'erreur global, vraisemblablement entre $\sim 0,1\%$ et $\sim 1,5\%$. Nous avons aussi formulé l'hypothèse que ces erreurs de comptage surviennent de manière uniformément aléatoire pour chaque bulletin de vote ce qui est plausible mais reste toutefois à vérifier.

En conclusion, ce travail nous a permis de modéliser la distribution des erreurs de comptage pour le cas de notre étude que sont les élections communales en Wallonie en 2018. Bien que nous manquons de données sur certaines hypothèses de notre modélisation, les résultats fournis par celle-ci nous paraissent

un bon indicateur sur la répartition des erreurs de comptage. En partant de cet indicateur, on peut alors calculer l'impact de ces erreurs sur les résultats des élections comme nous le faisons dans le post du blog decryptage.be.