

L'accusation de fraude est injustifiée

Michel Lasserre, le 31 juillet 2006

Les différentes analyses statistiques jusqu'à alors présentés (Husson-Coutrot, Duménil-Lévy, Chiche-Rounet) sont toutes fondées sur la même hypothèse de départ : les bulletins ont été classés par lots et par ordre alphabétique.

Cette première hypothèse repose pour sa part sur une deuxième hypothèse : le classement par ordre alphabétique implique un effet d'homogénéisation des différents lots.

Divers témoignages montrent que le classement alphabétique n'a été que partiellement respecté, ce qui suffit déjà à relativiser la valeur des conclusions que l'on peut tirer de ces différentes analyses. Toutefois, il semblerait que le classement ait bien été effectué au moins sur la première lettre des noms des votants. Il m'a alors paru judicieux de s'intéresser aux résultats globaux obtenus par les différentes lettres.

L'idée de départ était qu'un lot atypique doit normalement (en suivant la deuxième hypothèse) entraîner un effet sur l'ensemble des résultats de la lettre concernée. Par exemple, si la lettre X est partagée en 3 lots et que X3 monte un fort décalage sur certains candidats, ce décalage doit aussi se constater sur l'ensemble des bulletins X, c'est à dire : $X1 + X2 + X3$.

J'ai donc repris et complété le tableau fourni par l'huissier en calculant le pourcentage de voix obtenu pour chaque candidats. Ceci pour chaque lot, ainsi que pour chaque série de lots concernant la même première lettre. Je précise que certaines lettres consécutives sont cumulées du fait de l'existence de lots qui leurs sont communs (Ex : T et U car il existe un lot TU) . J'ai ensuite, à la fin du tableau (colonnes FG – FX), calculé l'écart de chaque candidat par rapport à ses résultats globaux en pourcentage (% du lot moins % global). Cet écart est "brut", c'est à dire non pondéré par rapport au poids de chaque lettre, ces calculs pourraient donc être améliorés mais je ne suis pas un spécialiste des statistiques et fait selon mes capacités. Une pondération ne changerait d'ailleurs pas la tendance observée.

Ce tableau est téléchargeable à cette adresse :

<http://m-lasserre.com/divers/tableau.xls>

Ce travail m'a permis de faire les observations suivantes :

1) Corrélation entre les lots atypiques et les regroupements par première lettre

La remise en cause de ces lots vient du fait qu'ils avantagent certains candidats (Coignard Valérie, Sabine Jauffret, Barral Audrey et Landfried Julien) au détriment d'autres (Coiffard-Grodoy Geneviève, Pradeau Raphaël, Coulomb Jean-Michel et Blasco Claudine), on s'intéressera donc plus particulièrement à leurs résultats.

a) On constate que pour les lots T3, L4, B3, B4, il existe bien une corrélation. Les écarts à la moyenne relevant de ces lots se retrouvent dans le total de chaque lettre, même si minorés par les autres lots de la même série. (Ex : le score de S. Jauffret à 64,04 % dans le lots T3 est en partie corrélé avec son score de 51,92 % dans le total de T). Les lots atypiques renvoient ici à des séries atypiques.

b) On constate que pour les lots P2, R2, P4, J1, il n'existe pas de corrélation. (S. Jauffret à 50,70 % en P2 n'est pas corrélé avec 38% en PQ).

Si la première constatation n'implique aucune remarque particulière, la deuxième pose questionnement. En observant l'ensemble des bulletins commençant par P et Q, on ne constate qu'un très faible écart avec les résultats globaux et cet ensemble n'est donc pas du tout atypique. Le lot atypique n'influe donc pas sur le total des bulletins commençant par la même lettre, il est compensé par les autres lots de la même série. Les lots atypiques ne renvoient pas ici à des séries atypiques. Attribuer l'atypie de ces lots à une fraude n'a alors guère de sens, car s'ils étaient "typiques" ce serait les séries correspondantes d'ensemble de bulletins qui ne le seraient pas.

La conclusion la plus probable est que ces lots n'ont rien de frauduleux mais que la répartition en lots n'a ici rien d'homogène.

On pourrait expliquer ce fait en s'appuyant sur un effet chronologique résultant d'un tri par ordre alphabétique défectueux, on verra dans la partie suivante qu'il existe aussi une autre explication.

2) Ordre alphabétique et homogénéité des séries

On s'intéressera ici aux colonnes de la fin du tableau et concernant les écarts entre le pourcentage global et celui de chaque série de bulletins commençant par la même lettre. L'hypothèse sur laquelle s'appuie les analyses est que le classement par ordre alphabétique permet d'obtenir des séries homogènes. Si cette hypothèse est confirmée, on doit donc s'attendre ici à trouver des séries présentant peu d'écart avec les résultats globaux. Qu'en est-il ?

Sur le tableau (colonnes FG – FX) j'ai surligné en violet les résultats individuels dont l'écart entre les pourcentages dépasse 9 points, et que l'on peut considérer comme de fortes divergences statistiques si les séries étaient censées être homogènes.

a) Si l'on observe les séries comportant des lots atypiques (B, G, IJK, L, PQ, R, TU, V), on constate que les série B, G, P, R et V ne contiennent aucune forte divergence, J et L en contiennent 1, et T 17.

b) Parmi les séries ne comportant pas de lots atypiques, A, C, D sont vierges de toute divergence, E en contient 11, F : 8, H : 16, M : 2, NO : 5, S : 6, WYZ : 24.

Ces constatations montre une absence d'homogénéité, en dehors de T les séries comprenant des lots atypiques montrent d'ailleurs curieusement plus d'homogénéité que les séries sans. **Ce qu'on peut en conclure est que le classement par ordre alphabétique ne permet pas ici d'obtenir des séries homogènes.**

La cause des importants écarts observés repose certainement sur la petite taille des lots dont les plus importants regroupent au maximum quelques centaines de bulletins. **Il est bien évident que si cette réalité concerne les séries, elle concerne aussi et à plus forte raison les lots qui contiennent encore moins de bulletins que les séries.**

Dans le contexte particulier de cette élection, où chacun est appelé à choisir 24 élus parmi 62 candidats, perturbée par un appel à voter pour une liste déterminée et dont tout le monde n'a pas eu connaissance, on peut s'interroger sur le choix de cette hypothèse d'homogénéité dans

une analyse statistique. Sur quelle base autre que purement théorique les analystes se fondent-ils pour avancer cette hypothèse ? A partir de quelle quantité de bulletins une telle hypothèse deviendrait-elle ici réaliste ? Plusieurs centaines de bulletins par lots, plusieurs milliers, plusieurs dizaines de milliers ? Pour répondre à cette question de manière scientifique il faudrait faire des études comparatives sur un grand nombre d'élections dont les données sont similaires à celles-ci, or de telles études n'existent certainement pas.

3) Vous avez dit fraude ?

L'accusation de fraude repose sur l'existence de lots atypiques, existence qui repose sur deux hypothèses de départ s'avérant erronées. **Sans le soutien de ces deux hypothèses, c'est la notion même de lots atypiques qui s'effondre, et l'accusation de fraude repose alors sur du vide.**

Si au lieu de s'intéresser à de simples lots les "fraudistes" avaient observé les séries, moins susceptibles d'écart que les lots, ils se seraient aperçu que parmi celles présentant le plus grand nombre d'écarts individuels (plus de 4) on en trouve une seule comprenant des lots "atypiques" (T) et six ne comprenant aucun de ces lots (E, F, H, NO, S, WYZ). On notera au passage que les séries E F H S, fortes en anomalies, ont été entièrement dépouillées les 10 et 11 Juin, soit avant la date avancée par les "fraudistes" pour justifier leur accusation.

Le lot T3, qui regroupe 114 bulletins, montre des écarts plus importants que les autres lots, et est mis en avant par les "fraudistes" pour justifier leur thèse. On a déjà observé que ses écarts sont largement relativisés sur l'ensemble de la série TU. Face à l'hétérogénéité des séries, il n'y a rien d'extraordinaire à envisager que sa haute teneur en bulletins défavorables aux candidats de la liste Susan Georges soit le résultat d'un simple hasard, rien de plus extraordinaire qu'à envisager qu'il soit le résultat d'une fraude.

4) Pour conclure.

Cette étude montre que l'accusation de fraude s'appuie sur des interprétations statistiques fondées sur des hypothèses erronées.

Une fraude est grave, une accusation de fraude non justifiée l'est tout autant sinon plus. Mettre publiquement en accusation des camarades, profondément perturber le fonctionnement de l'association en refusant de siéger au CA, sont deux pratiques qui pourraient se comprendre si l'on avait des preuves d'une fraude. Sans ces preuves, ces deux pratiques sont totalement inacceptables. En lançant ainsi la rumeur d'une fraude, et en agissant comme si cette fraude était avérée, les "fraudistes" mettent en grave péril l'association. Ce n'est certainement pas leur objectif, mais c'est pourtant une réalité sur laquelle il serait fortement souhaitable qu'ils s'interrogent le plus rapidement possible.