

Anatomie statistique d'une fraude

Michel Husson, 4 juillet 2006

Ce texte synthétise plusieurs contributions antérieures et bénéficie des nombreuses réactions qu'elles ont pu susciter (voir <http://hussonet.free.fr/attac0.htm>). Les considérations plus personnelles qui figuraient en introduction à la première version de cette analyse ([Radiographie d'une blessure](#)) ne sont pas reproduites ici, afin de conserver à ce document son caractère d'analyse. Cette analyse sera menée autour de ces deux principales questions :

- 1) **une partie des groupes de bulletins présentent-ils des anomalies statistiques ?**
- 2) **ces anomalies peuvent-elles être expliquées autrement que par une fraude ?**

1. Le repérage des anomalies statistiques

Le vote pour l'élection des représentants des membres actifs ne s'est pas déroulé selon le protocole prévu au départ (<http://hussonet.free.fr/procedur.pdf>). Les bulletins ont bien été classés par ordre alphabétique en fonction du nom des votants, mais ils n'ont pas été dépouillés au fur et à mesure que les enveloppes qui les contenaient étaient décachetées. Des récapitulatifs intermédiaires ont été établis, et leur résultat connu en temps réel. Ils ont permis de déceler rapidement des anomalies statistiques, corroborées par les observations de plusieurs scrutateurs.

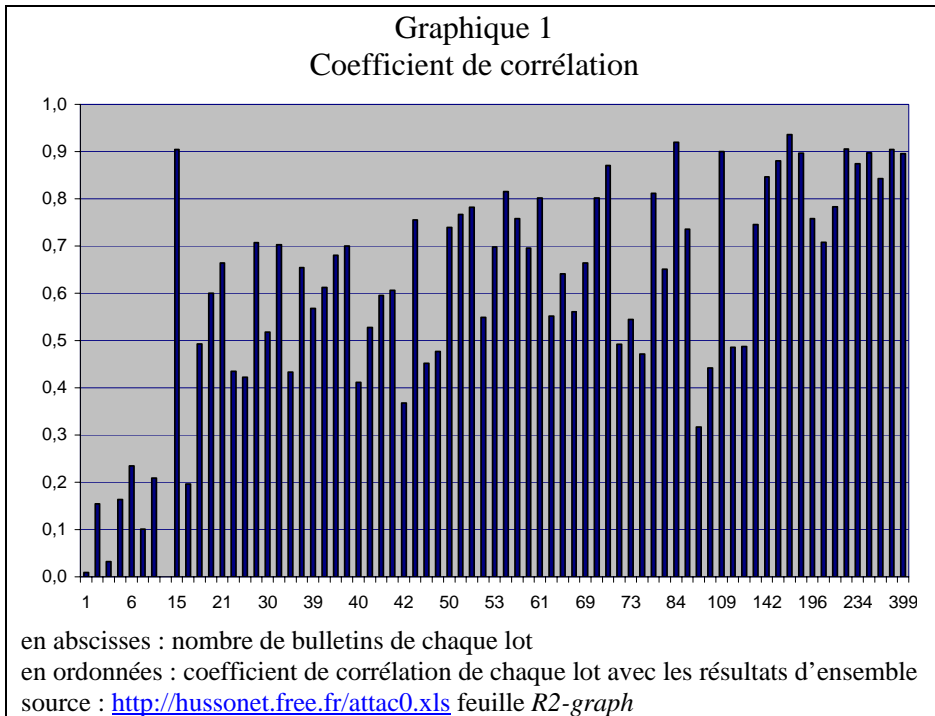
Pour systématiser ces premiers résultats, nous disposons, après recomptage par huissier, d'un tableau de résultat informatique donnant (en ligne) le nombre de voix obtenu par chaque candidat, ventilé (en colonnes) par « lots » (<http://hussonet.free.fr/huissier.xls>). Chaque lettre de l'alphabet correspond ainsi à un ou plusieurs lots. On connaît aussi la chronologie de dépouillement, qui ne suit pas l'ordre alphabétique, et on dispose de récapitulatifs intermédiaires.

1.1. La méthode de repérage

Le résultat de chaque lot a été comparé au résultat d'ensemble. On a calculé à chaque fois le **coefficient de corrélation** entre chaque lot et les résultats. Rappelons que ce coefficient varie entre 0 et 1 : plus il est élevé, plus les deux distributions comparées sont voisines. Ce test est adéquat au problème posé, et nous l'avons préféré au test du Khi-deux (pour plus de détails voir <http://hussonet.free.fr/co2.pdf>).

Le graphique 1 donne pour chacun des lots, préalablement classés selon le nombre de bulletins qu'ils contiennent, leur coefficient de corrélation par rapport aux résultats d'ensemble. La question de la taille est ici primordiale puisqu'un lot donné pourra d'autant plus s'écarter de la distribution d'ensemble qu'il est petit.

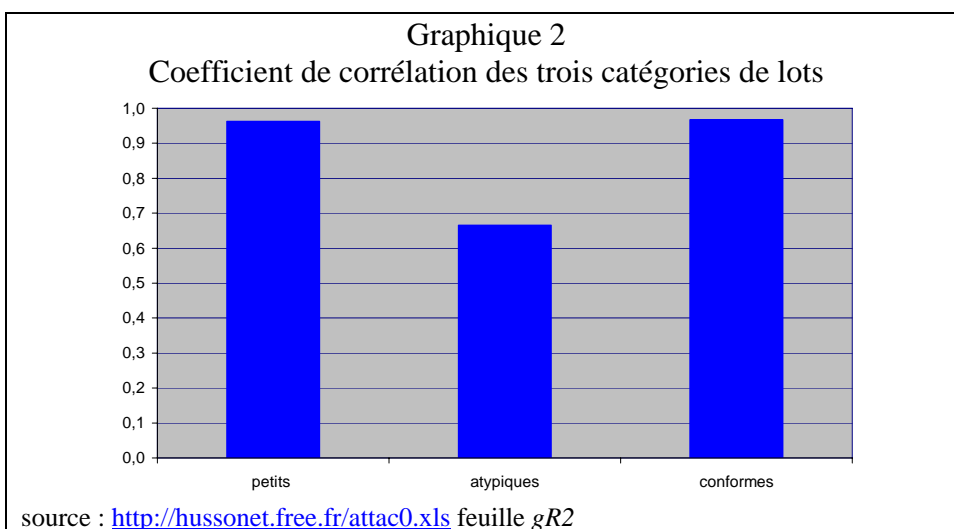
On vérifie bien que les lots les plus gros ont des coefficients de corrélation supérieurs. Mais cette liaison n'est cependant pas uniforme. A gauche du graphique, les lots de petite taille ont des coefficients de corrélation très variables, mais certains sont presque aussi bons que ceux de grande taille. Ce constat est parfaitement conforme à la théorie statistique : les petits lots ayant une probabilité plus grande de s'écarter des résultats globaux, certains seront très mauvais mais d'autres s'approcheront du résultat global ou même le reproduiront.



En revanche, le graphique présente une anomalie : certains lots de taille moyenne (autour de 100 bulletins) ont des coefficients de corrélation particulièrement bas. Mais il ne s'agit que d'une intuition visuelle, qui doit être vérifiée. Pour cela, nous avons commencé par construire trois catégories de lots :

- lots **petits** : moins de 50 bulletins ;
- lots **atypiques** : plus de 50 bulletins mais coefficient de corrélation inférieur à 0,65 ;
- lots **conformes** : plus de 50 bulletins et coefficient de corrélation supérieur à 0,65.

Nous caractérisons a priori la catégorie intermédiaire d'atypique, parce qu'il s'agit de lots très distincts de la distribution d'ensemble, en dépit d'une taille suffisante. Pour vérifier la pertinence de ces catégories *a priori*, nous avons additionné les lots composant chacune de ces catégories et avons à nouveau comparé la distribution obtenue à celle de l'ensemble.



On voit alors apparaître (graphique 2) un premier résultat robuste détaillé dans le tableau : **regroupés en un seul, les 35 petits lots (soit 1000 bulletins) font apparaître un coefficient de corrélation de 0,962** très voisin de celui (0,967) qu’obtiennent, une fois regroupés de la même manière, les 27 lots jugés a priori conformes. C’est pourquoi on les considérera dorénavant comme conformes.

En revanche, **les 11 lots considérés a priori comme atypiques confirment qu’ils le sont effectivement**, puisque cette catégorie une fois agrégée continue à obtenir un coefficient de corrélation significativement inférieur (0,665).

Tableau 1

Coefficients de corrélation avec le résultat global par catégories de lots

catégorie de lots	nombre de lots	nombre de bulletins	coefficient de corrélation avec le Segment 1
petits	35	1000	0,962
atypiques	11	910	0,665
conformes	27	3938	0,967
total	73	5848	-

source : <http://hussonet.free.fr/attac0.xls> feuille R2

1.2. Les caractéristiques des onze lots atypiques

Le point important à souligner en préalable est que **tous les lots atypiques sont postérieurs au récapitulatif en date du 11 juin** qui porte sur 2658 bulletins, soit 45 % (tableau 2).

Tableau 2

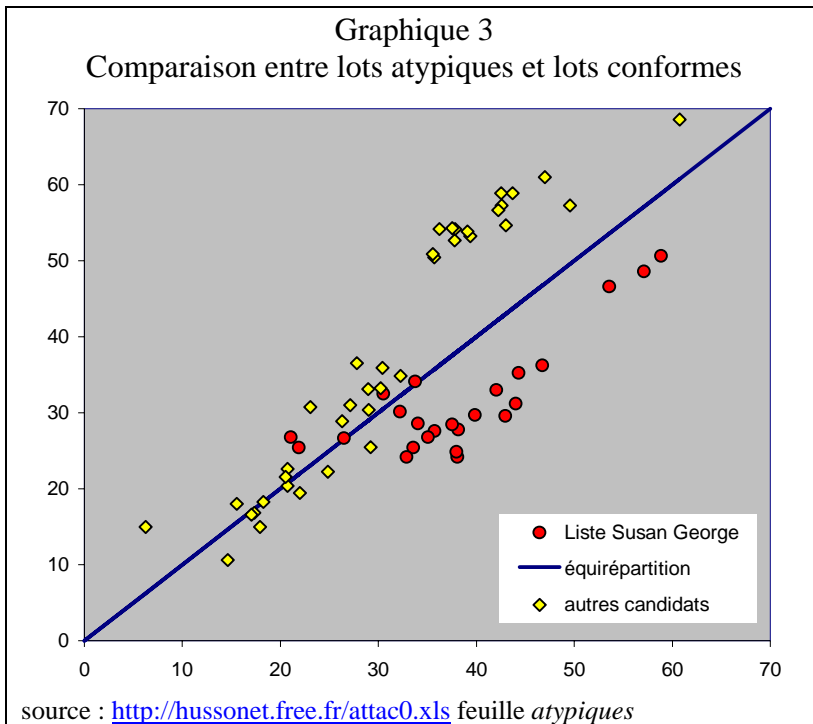
Caractéristiques des onze lots atypiques

date du lot	lot	nombre de bulletins	coefficient de corrélation
cumul 12 Juin	D4	67	0,561
cumul 12 Juin	G4	51	0,549
cumul 13 Juin	G3	65	0,641
cumul 14 Juin	B3	100	0,442
cumul 14 Juin	B4	100	0,317
cumul 14 Juin	L4	110	0,486
cumul 15 Juin	C1	73	0,545
cumul 15 Juin	P2	71	0,492
cumul 15 Juin	R2	75	0,472
cumul 15 Juin	T3	114	0,487
cumul 15 Juin	V1	84	0,650
Total		910	0,665

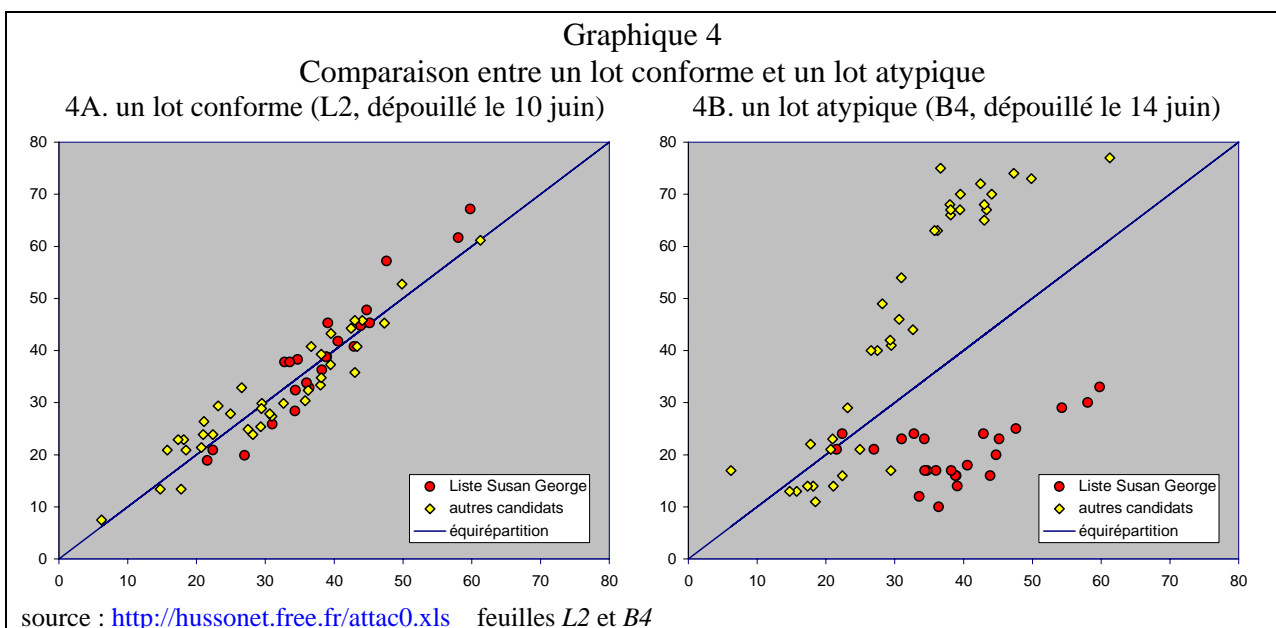
source : <http://hussonet.free.fr/attac0.xls> feuille *cumulu*

Tous les lots atypiques présentent la même configuration : les candidats de la liste Susan George (voir sa lettre : <http://hussonet.free.fr/sgeorge.pdf>) obtiennent moins de voix que dans les lots conformes alors que certains des autres candidats obtiennent plus.

Cette distorsion apparaît clairement dans le graphique 3 qui compare la répartition des scores dans les lots atypiques à celle des lots conformes.



On retrouve cette configuration pour chacun des onze lots atypiques, de façon d'autant plus marquée que le coefficient de corrélation est faible. Une telle configuration s'oppose nettement à celle des lots conformes (graphique 4).



2. Des anomalies statistiques à la fraude

2.1. Impact des lots atypiques sur le résultat final

Il faut au préalable vérifier dans quelle mesure les lots atypiques ont influé sur le résultat final. La méthode retenue est simple : nous attribuons à chaque candidat le pourcentage de voix qu'il aurait obtenu en ne prenant en considération que les seuls lots conformes, puis nous appliquons ce pourcentage au nombre de bulletins dépouillés. Nous établissons la composition du CA qui en aurait découlé, et nous le comparons à la composition qui résulte du recomptage de l'huissier.

Il faut signaler au passage que cette composition diffère à la marge des résultats annoncés lors de l'AG de Rennes (<http://hussonet.free.fr/atta6454.pdf>) : Michel Gicquel, élu en 24ème position avec 2291 voix (37,86%) n'en obtient plus que 2232 et passe donc derrière Julien Landfried (2233 voix) qui devrait donc siéger au CA à sa place.

Tableau 3

Composition du CA avec ou sans les lots atypiques

Composition effective				Composition sur la base des lots conformes			
rang	Nom	voix	%	rang	nom	voix	%
1	NIKONOFF Jacques	3651	62,4	1	NIKONOFF Jacques	3584	61,3
2	TROUVE Aurélie	3384	57,9	2	TROUVE Aurélie	3496	59,8
3	HARRIBEY Jean-Marie	3280	56,1	3	HARRIBEY Jean-Marie	3397	58,1
4	AZAM Geneviève	3087	52,8	4	AZAM Geneviève	3177	54,3
5	TASSI Régine	2990	51,1	5	TASSI Régine	2917	49,9
6	JONQUET Bernadette	2891	49,4	6	VEILHAN Bénédicte	2784	47,6
7	KARBOWSKA Monika	2707	46,3	7	JONQUET Bernadette	2768	47,3
8	VEILHAN Bénédicte	2663	45,5	8	BÉNÉTEAU Chloé	2642	45,2
9	WEBER Jacques	2651	45,3	9	BAUNEZ Christelle	2618	44,8
10	BERNIER Aurélien	2640	45,1	10	KARBOWSKA Monika	2581	44,1
11	LUDI Aurore	2636	45,1	11	ROLET Lysiane	2568	43,9
12	GOUSSOT Danielle	2615	44,7	12	BERNIER Aurélien	2534	43,3
13	BÉNÉTEAU Chloé	2531	43,3	13	LUDI Aurore	2515	43,0
14	BAUNEZ Christelle	2487	42,5	14	WEBER Jacques	2514	43,0
15	LAMBERT Renaud	2440	41,7	15	VERDIN Cécile	2510	42,9
16	LAURENT Matthieu	2433	41,6	16	GOUSSOT Danielle	2485	42,5
17	ROLET Lysiane	2421	41,4	17	MERCIER Isabelle	2373	40,6
18	VERDIN Cécile	2401	41,1	18	LAMBERT Renaud	2312	39,5
19	GAZIELLO Emmanuelle	2372	40,6	19	LAURENT Matthieu	2311	39,5
20	COIGNARD Valérie	2361	40,4	20	COIFFARD-GROSDOY Geneviève	2286	39,1
21	BARRAL Audrey	2358	40,3	21	PRADEAU Raphael	2279	39,0
22	JAUFFRET Sabine	2294	39,2	22	COULOMB Jean Michel	2270	38,8
23	MERCIER Isabelle	2265	38,7	23	BLASCO Claudine	2235	38,2
24	LANDFRIED Julien	2233	38,2	24	GAZIELLO Emmanuelle	2230	38,1

En couleur, les candidat(e)s figurant sur la « liste Susan George »

source : <http://hussonet.free.fr/attac0.xls> feuille VOTE

Les résultats de cette évaluation figurent dans le tableau 3. Ils montrent que si les seuls lots conformes avaient été pris en compte, la liste de Susan George aurait obtenu 13 représentants au CA sur 24 sièges, alors qu'elle n'en a eu que 9 sur 24 (Régine Tassi n'est pas comptée comme faisant partie de la liste Susan George).

Les lots atypiques conduisent donc à la permutation de quatre candidats : Geneviève Coiffard-Grosdoy, Raphaël Pradeau, Jean-Michel Coulomb et Claudine Blasco auraient été élus au CA en lieu et place de Valérie Coignard, Audrey Barral, Sabine Jauffret et Julien Landfried (ou Michel Gicquel).

Le vote était très serré au « milieu du peloton », de sorte que les transferts de voix induits par les lots atypiques ont pu modifier profondément la physionomie du CA, alors même qu'ils n'ont porté au maximum que sur 150 bulletins. Certes, les lots atypiques en regroupent un nombre de bulletins plus élevés (910) mais le redressement ne porte que sur les bulletins en plus ou en moins de la référence des lots conformes (voir tableau 4).

Tableau 4
Transferts de voix engendrés par les lots atypiques

	% final	% hors atypiques	différence en %	différence en voix
Quatre candidats de la liste Susan George « sortent »...				
COIFFARD-GROSDOY Geneviève	37,0	39,1	- 2,1	- 120
PRADEAU Raphaël	36,5	39,0	- 2,5	-146
COULOMB Jean-Michel	36,5	38,8	- 2,3	-137
BLASCO Claudine	36,5	38,2	- 1,7	- 102
... et sont remplacés par quatre autres				
COIGNARD Valérie	40,4	38,0	+2,4	+138
BARRAL Audrey	40,3	38,1	+2,2	+128
JAUFFRET Sabine	39,2	36,6	+2,6	+152
LANDFRIED Julien	38,2	36,2	+2,0	+115

2.2. Les explications alternatives à la fraude

Après avoir établi qu'il existe un sous-ensemble de bulletins caractérisé par des anomalies non aléatoires, il convient d'examiner si elles peuvent s'expliquer autrement que par la fraude. Deux effets possibles sont envisageables : effet de liste et/ou effet de génération. Dans les deux cas, ces schémas explicatifs rejettent l'idée selon laquelle la ventilation des bulletins par ordre alphabétique suffirait à garantir une distribution aléatoire des bulletins.

L'effet de liste consiste à dire que l'annonce d'une liste par Susan George, suivie d'une déclaration signée par 32 candidats qui récusait cette démarche (mais constituaient de fait une contre-liste), a pu influencer sur les bulletins non encore envoyés. Les bulletins reçus en dernier seraient donc, plus que les premiers, sensibles à cette logique de liste. Par ailleurs, des adhérents auraient pu être sollicités dans les derniers jours, et émettre des bulletins fortement typés dans un sens ou dans l'autre. Enfin, des groupes d'adhérents auraient pu se concerter et envoyer des bulletins semblables (*effet de grappe*). Ces effets conduiraient à une apparence d'anomalie mais qui ne ferait que refléter des comportements licites n'enfreignant donc pas la sincérité du vote.

L'effet de grappe (plusieurs adhérents votent de manière identique) ne peut pourtant expliquer à lui seul la configuration d'ensemble, parce qu'il aurait été neutralisé par le tri alphabétique. Imaginons que les membres d'un comité émettent un vote-type décidé en commun. Ce tir groupé sera en fait dispersé, puisque le bulletin de Durand se retrouvera à la lettre D, et celui de Martin à la lettre M. Si cet effet devait néanmoins apparaître, il aurait dû se manifester de manière relativement uniforme entre les lots (de telle sorte qu'il serait d'ailleurs indécélable). Enfin, il faut rappeler que la plupart des lettres de l'alphabet sont ventilées sur plusieurs lots, dont certains seulement sont atypiques. Conclusion : l'effet de grappe n'est pas à lui seul un schéma adéquat à la configuration observée.

Il faut en réalité imaginer un croisement des deux effets : il y a des votes fortement typés et ils n'apparaissent que dans les bulletins reçus en dernier. Cela permettrait d'expliquer pourquoi les lots atypiques n'apparaissent qu'après le second récapitulatif en date du 11 Juin (45 % des bulletins reçus). Ce résultat à mi-parcours (baptisé Segment 1 dans d'autres analyses) regrouperait donc des bulletins encore peu marqués par la logique de liste qui n'apparaîtrait que sur la seconde moitié des bulletins reçus plus tardivement.

Cette construction repose sur un postulat selon lequel les bulletins ont été dépouillés dans l'ordre d'arrivée. Mais, même en admettant ce postulat, l'effet invoqué ne peut rendre compte de la ventilation dans le temps des lots atypiques. Ainsi, la lettre D est conforme pour la partie dépouillée le 11 juin, puis atypique le 12, puis de nouveau conforme les 13, 14 et 15. La lettre G est conforme le 11, puis atypique le 12 et le 13, et de nouveau conforme le 15. La fin des lettres P, R, T et V a été dépouillée en deux lots en date du 15 juin : pour chacune de ces lettres, on a un lot atypique et un lot conforme. Enfin, d'autres lettres sont conformes de bout en bout.

Bref, si le dépouillement est chronologique (il commence par les premiers bulletins reçus) cette structure ne confirme pas un effet de grappe combiné à une effet de date. Dans ce cas en effet, les lots atypiques devraient se retrouver dans les dépouillements les plus tardifs. Certes, ils n'apparaissent qu'après le 11 Juin mais, à partir de cette date, leur répartition dans le temps se fait de manière très différente d'une lettre à l'autre. Et si le postulat chronologique ne correspond pas au déroulement concret du dépouillement, cela veut dire que les lots ont été dépouillés de manière aléatoire : dans ce cas, l'effet en question devrait se manifester sur tous les lots, de telle sorte qu'il ne serait pas identifiable. Bref, **les schémas alternatifs à la fraude ne rendent pas compte de la configuration observée.**

2.4. L'hypothèse d'une « contre-fraude »

Pour être complet, il faut traiter cette question : qui sont les « atypiques » et qui sont les « conformes » ? Poser cette question, c'est suggérer que la fraude pourrait aussi bien provenir de la liste de Susan George. Quatre arguments permettent de rejeter cette ligne d'interprétation.

a) la proportion : d'un côté, 910 bulletins, de l'autre 4938 bulletins soit 5 fois plus. Il aurait fallu une fraude extraordinairement massive (pour ne pas arriver au résultat souhaité).

b) la chronologie : les lots atypiques ont été dépouillés **après** la première moitié des bulletins et ont été suivis de nouveaux lots redonnant les mêmes résultats que les 3000 premiers. On voit mal comment et pourquoi la fraude aurait porté d'une manière extraordinairement bien répartie sur les 3000 premiers bulletins, puis aurait cessé au moment où elle était la plus facile à mettre en œuvre,

pour reprendre ensuite. Pour que cette fraude ne soit pas décelable statistiquement, il aurait fallu en outre qu'elle porte sur l'intégralité des lots, avec une très grande régularité.

c) le résultat final : il faut imaginer des scénarios particulièrement compliqués pour penser que les fraudeurs, capables de manipuler 85 % des lots représentant 84 % des bulletins, n'auraient pas poussé l'avantage jusqu'à faire basculer le score final en leur faveur.

d) il reste un quatrième élément qu'il sera peut-être possible d'analyser : c'est la structure interne des lots atypiques qui pourrait permettre de repérer plus finement la présence de bulletins individuels très voisins, voire identiques. Il y a là une voie qui permettrait de mieux éclairer encore le constat d'ensemble.

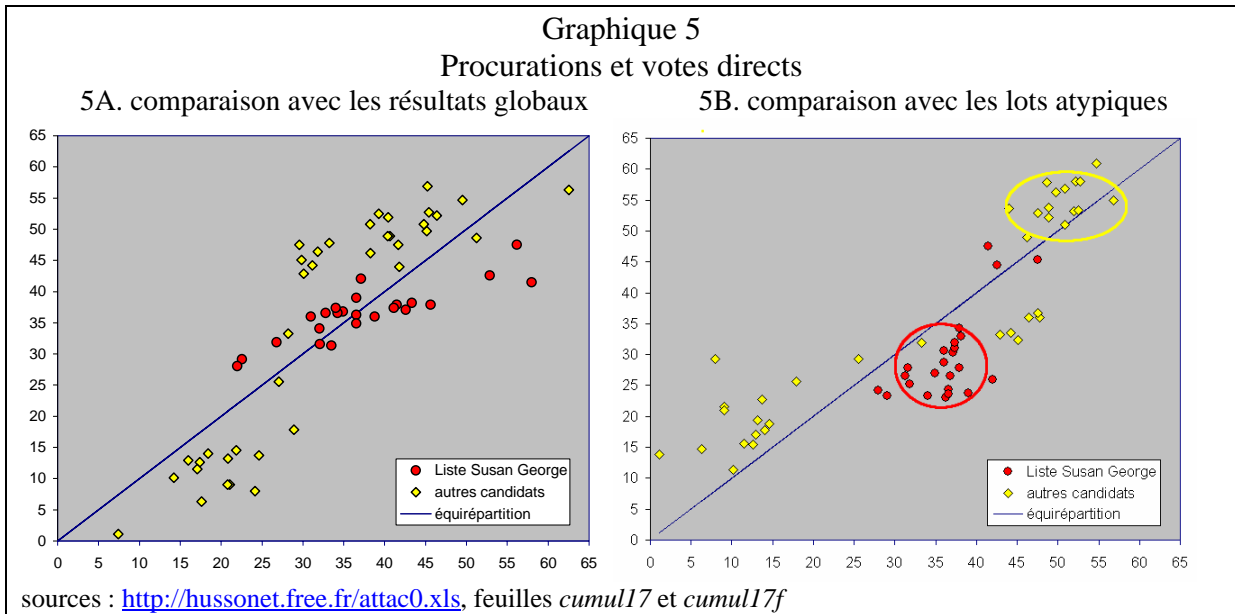
2.5. La structure spécifique des anomalies est la marque d'une fraude

On a déjà constaté que les lots atypiques défavorisent **les candidat(e)s de la « liste Susan George selon un ciblage précis** et rendent peu plausible un phénomène aléatoire.

L'hypothèse de la fraude peut alors être soumise au test suivant : les anomalies constatées sont-elles distribuées de manière indépendante ou non par rapport aux résultats intermédiaires fournis par le cumul du 11 Juin. Si la distribution des anomalies est indépendante de l'information que livrent ces résultats à mi-parcours, alors la fraude n'est pas établie. Si au contraire elle en dépend, alors la fraude est établie. Tout schéma explicatif qui renvoie à un comportement licite doit en effet être rejeté s'il est établi que les bulletins atypiques intègrent une telle information : en effet, **aucun bulletin licite n'a été établi après la publication restreinte des résultats intermédiaires et n'a donc pu intégrer l'information qu'ils donnaient.**

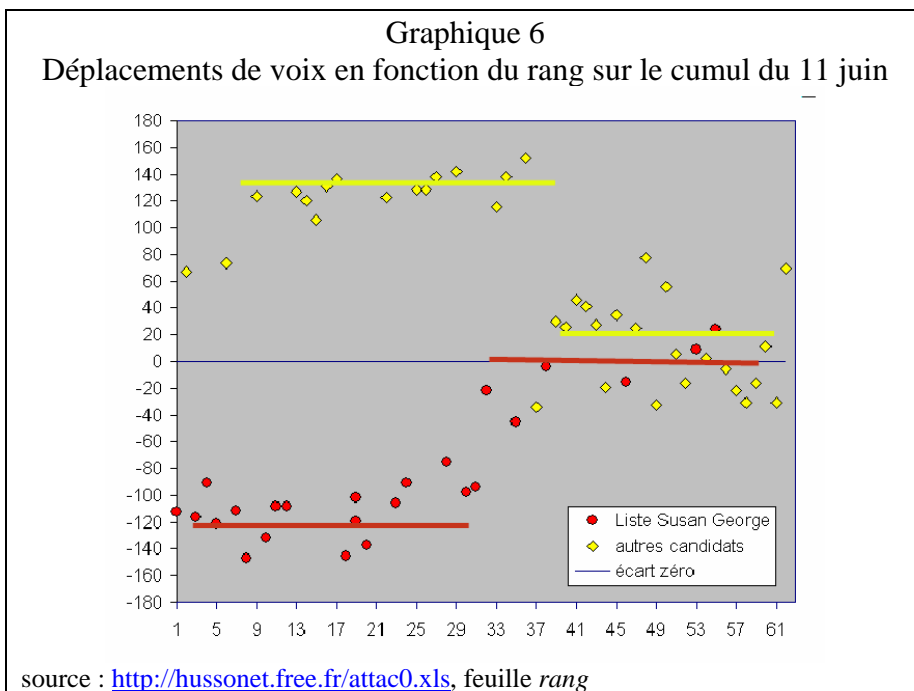
La seule exception envisageable à ce principe est celle des votes **directs et par procuration**. 281 procurations et 83 votes directs collectés lors de l'AG de Rennes ont été ainsi regroupés dans le cumul du 17 Juin. On peut considérer ces 364 bulletins comme particulièrement sensibles à des effets de liste. Si l'on suppose que les lots atypiques ne font que refléter de tels effets, leur structure particulière devrait se retrouver, de manière encore plus marquée, parmi les votes directs et par procuration.

Or, ce n'est pas vraiment le cas : par rapport aux résultats d'ensemble, la liste Susan George retrouve des résultats à peu près semblables. Les votes en faveur des autres candidats font apparaître une structure particulière de résultats : tout se passe comme si les partisans de cette liste commettaient une erreur tactique en votant moins que la moyenne pour les candidats qui seront finalement les mieux placés, et en donnant un coup de pouce à ceux qui seront finalement éliminés (graphique 5A). Ce point est décisif : il montre que ces votes n'intègrent pas l'information fournie par le dépouillement à mi-parcours.



Même par rapport aux votes directs et par procuration, les lots atypiques restent donc nettement plus discriminants, avec une baisse ciblée pour la liste Susan George et une hausse tout aussi ciblée pour les mieux placés des autres candidats (graphique 5B).

On en arrive alors à la dernière étape qui permet de véritablement tester l’hypothèse de la fraude. Elle consiste à examiner la dépendance des transferts de voix associés aux lots atypiques aux résultats intermédiaires du 11 Juin. Le graphique 6 répond à cette question. Il montre que la distribution des déplacements de voix entre lots atypiques et lots conformes dépend du rang de chaque candidat à mi-parcours : **les candidats en position éligible bénéficient ou pâtissent de déplacements de voix plus élevés** qui varient à peu près entre 100 et 140 voix.



Puis, au-delà du 30ème rang, les déplacements baissent assez nettement, parce que les candidats concernés sont de toute manière hors course. On tient là la preuve de la fraude : **les résultats globaux ne sont pas indépendants d'une information ignorée des adhérents au moment où ils remplissaient leur bulletin, mais dont disposaient les personnes présentes sur le lieu du dépouillement.**

Les résultats du vote ne présentent pas seulement des « anomalies » mais sont biaisés en fonction d'une intention délibérée qui ne doit rien au hasard : il s'agissait de les infléchir au détriment de la liste Susan George. L'information disponible ne portait pas seulement sur le rapport de forces global à mi-parcours (14 sièges sur 24 à la liste de Susan George) mais aussi sur la liste nominative des candidats dont le score devait être augmenté ou baisser pour que la composition du CA soit modifiée dans le sens souhaité. Les résultats de ce vote sont donc marqués du signe de la fraude.