

# Un examen statistique des résultats du scrutin pour les élections des représentants "membres actifs" au Conseil d'administration d'Attac

Danièle Guillemot, Laurent Davezies, Loup Wolff (statisticiens économistes)

18/08/06

Cette étude a été menée à partir du fichier, établi sous contrôle d'huissier, qui nous a été transmis et qui récapitule les résultats du dépouillement des bulletins de vote par correspondance. Ce fichier se présente sous la forme d'un tableau faisant figurer en ligne les noms des candidats et en colonne les différents "lots" dans lesquels les bulletins de vote avaient été préalablement regroupés. Chacun de ces lots a été constitué en regroupant les bulletins de vote des électeurs en fonction de la première lettre de leur nom et a été dépouillé sans que l'ordre d'arrivée des paquets ne soit strictement respecté.

Chaque cellule de ce tableau donne donc le nombre de voix exprimées en faveur d'un candidat – en ligne – dans un lot donné – en colonne.

## L'expérience

L'objectif de cette étude est de repérer les cellules du tableau qui s'écartent le plus d'une certaine "régularité statistique" qui voudrait que les lignes et les colonnes du tableau transmis soient largement indépendantes : la première lettre du nom de famille des votants devrait avoir une influence faible sur leurs comportements de vote.

Ce ne seront pas tant les cellules irrégulières, en elles-mêmes, qui attireront notre attention (il est tout à fait normal dans toute expérience réelle d'être confronté à des irrégularités statistiques). Mais plutôt une certaine *régularité de ces irrégularités* en faveur ou en défaveur de certains candidats et relativement à la chronologie du dépouillement des lots de bulletin.

Nous laisserons aux futurs lecteurs de cette note le soin de juger de la nature suspecte ou non de ces régularités, en fonction de leur connaissance de l'association et de ses membres.

Nous voudrions souligner ici notre regret de n'avoir pu travailler sur les bulletins de vote eux-mêmes qui, en nous permettant d'identifier les comportements de vote individuels, nous auraient permis de mieux prendre en compte les modalités de ce scrutin et de chiffrer le degré de probabilité des écarts observés.

## Le principe

Nous nous sommes servis, pour la quantification du degré d'irrégularité statistique des cellules, de la statistique du *Chi 2*. Il s'agit d'une statistique qui, pour chaque candidat  $i$  (en ligne) et pour chaque lot  $j$  (en colonne), compare le nombre de votes exprimés *observé* ( $n_{ij}$ ) au nombre théorique de votes qui devrait s'afficher dans cette cellule, si l'hypothèse de

l'indépendance entre la première lettre du nom de famille des votants et les comportements de vote était vérifiée.

Nous avons donc calculé pour chaque cellule la valeur du Chi 2 correspondante, ajustée sur les marges (i.e. en rapport avec le nombre de voies comptabilisées dans le lot  $n_i$  et le nombre de voies enregistré pour le candidat tous lots confondus  $n_j$ ). Cette statistique vaut 0 si  $n_{ij} = (n_i * n_j)/N$  (où N est le nombre total de votes exprimés) et prend une valeur positive si  $n_{ij}$  s'écarte du deuxième membre de l'égalité. Les valeurs positives du test pointent donc sur des cellules "irrégulières" dans le tableau. Et la cellule est d'autant plus irrégulière que la valeur du Chi 2 correspondante est élevée.

Il est de plus à noter que l'expérience n'a été menée que sur les plus gros lots (i.e. comptabilisant plus de 70 votes exprimés). La statistique du Chi 2 ne fonctionne correctement que lorsque suffisamment d'individus sont concernés. Les lots étudiés sont les suivants : A3, L2, C5, D5, F1, G2, H1, M1, P3, R1, S1, V3, C3, C2, D2, L3, B2, B3, B4, L4, B1, C1, P2, R2, T3, V1.

En l'absence du détail des bulletins de vote dépouillés, ce type de statistiques ne nous permettront pas de calculer des tests sur les distributions par candidats ou par lots ou de chiffrer le degré de probabilité des écarts observés. Il peut être plus simplement utilisé pour repérer des écarts.

### Résumé des irrégularités observées

Le tableau qui suit présente la liste des 100 cellules les plus irrégulières. On peut y lire la valeur du Chi2 ainsi que, colonne suivante, le sens de l'irrégularité observée : si la colonne Sens présente le signe "-", c'est que le nombre de votes y est anormalement bas et inversement si c'est le signe "+". La valeur indiquée dans la colonne P-Value donne la probabilité que la statistique de test soit supérieure ou égale à la valeur observée sous l'hypothèse que le mode de regroupement des bulletins par lot pour le dépouillement n'induit pas de sur-représentation. Par exemple sous cette hypothèse, Geneviève Azam n'avait que 1.26 chance sur 10000 d'obtenir un aussi faible score dans le lot B3.

Nom	Lots	Valeur du test	P-Value	Sens
AZAM Geneviève	B3	14,7	1,26E-04	-
	B4	23,3	1,39E-06	-
	L2	123,9	8,86E-29	-
	L4	18,4	1,79E-05	-
BENETEAU Chloé	B4	16,8	4,15E-05	-
BARRAL Audrey	B3	14,5	1,40E-04	+
	B4	29,5	5,59E-08	+
	L4	19,3	1,12E-05	+
	T3	16,1	6,01E-05	+
BAUNEZ Christelle	B4	20,6	5,66E-06	-

	L4	14,4	1,48E-04	-
	T3	18	2,21E-05	-
BENOIT Marie Louise	B4	19,7	9,06E-06	+
BERNIER Aurélien	B4	21,5	3,54E-06	+
BLASCO Claudine	B4	16,4	5,13E-05	-
BOIS Alain	B3	19,7	9,06E-06	+
	C1	61,7	4,00E-15	+
	P2	16,7	4,38E-05	+
BRUSTIER Gaël	B4	24,4	7,83E-07	+
CLOCHEPIN Philippe	B4	17,3	3,19E-05	-
	C1	22,6	1,99E-06	-
COIFFARD GROSDOY Geneviève	B4	21,5	3,54E-06	-
	T3	14,2	1,64E-04	-
COIGNARD Valérie	B3	26,3	2,92E-07	+
	B4	30,7	3,01E-08	+
	R2	14,9	1,13E-04	+
	T3	27,5	1,57E-07	+
COULOMB Jean Michel	B4	17,5	2,87E-05	-
	T3	28	1,21E-07	-
DENIMAL Florent	B3	22,3	2,33E-06	+
DUMAS Renaud	B3	14,5	1,40E-04	+
ETCHEGARAY José	L4	16	6,33E-05	-
GAYOSO José	B2	17,1	3,55E-05	+
GAZIELLO Emmanuelle	B3	19,1	1,24E-05	+
	B4	26,8	2,26E-07	+
	G2	15,4	8,70E-05	-
	L4	17,2	3,36E-05	+
	T3	22,3	2,33E-06	+
GICQUEL Michel	B3	21,3	3,93E-06	+
	B4	27,4	1,65E-07	+
	L4	19,8	8,60E-06	+
	T3	18,3	1,89E-05	+
GOUSSOT Danielle	B3	16,7	4,38E-05	+
	B4	30,2	3,90E-08	+
	T3	18,9	1,38E-05	+
HARRIBEY Jean Marie	B4	29,9	4,55E-08	-

	R2	15,4	8,70E-05	-
JAUFFRET Sabine	B3	16,4	5,13E-05	+
	B4	53,6	2,46E-13	+
	C1	14,1	1,73E-04	-
	H1	13,8	2,03E-04	-
	L4	31,7	1,80E-08	+
	R2	15,3	9,17E-05	+
	T3	29,4	5,89E-08	+
JONQUET Bernadette	B4	23,9	1,01E-06	+
	L4	15,3	9,17E-05	+
	T3	16,2	5,70E-05	+
JOUVE Gérard	B3	23,2	1,46E-06	+
KARBOWSKA Monika	B4	22,6	1,99E-06	+
	L4	16,3	5,41E-05	+
	T3	26,2	3,08E-07	+
LAMBERT Renaud	B4	33	9,22E-09	+
	H1	15,1	1,02E-04	-
	T3	16,7	4,38E-05	+
	V1	14,1	1,73E-04	+
LANDFRIED Julien	B3	14,9	1,13E-04	+
	B4	26	3,41E-07	+
	L4	15,3	9,17E-05	+
LAURENT Matthieu	B4	27,3	1,74E-07	+
	T3	19	1,31E-05	+
LUDI Aurore	B4	19,7	9,06E-06	+
	L4	15,4	8,70E-05	+
	T3	15,1	1,02E-04	+
MERCIER Isabelle	B4	17,7	2,59E-05	-
	T3	14,6	1,33E-04	-
NURIER Roland	B3	26,7	2,38E-07	+
	C1	17,1	3,55E-05	+
PRADEAU Raphael	B4	17,8	2,45E-05	-
	L4	15,6	7,83E-05	-
	T3	14,2	1,64E-04	-
ROLET Lysiane	B4	27,2	1,83E-07	-
	R1	19,4	1,06E-05	+

	T3	18,4	1,79E-05	-
ROUQUET Thierry	B4	26,1	3,24E-07	-
TASSI Régine	B4	18,4	1,79E-05	+
TOSTI Jean	B3	15,5	8,25E-05	+
TRIC Olivier	C1	14,8	1,20E-04	+
TROUVE Aurélie	B4	29	7,24E-08	-
	R2	25,7	3,99E-07	-
	T3	17,2	3,36E-05	-
VEILHAN Bénédicte	B4	17,6	2,73E-05	-
	L4	16,9	3,94E-05	-
	M1	22,9	1,71E-06	+
	T3	17,6	2,73E-05	-
VERDIN Cécile	T3	14,4	1,48E-04	-
VIALE Frédéric	P2	21,6	3,36E-06	+
WEBER Jacques	B3	19,9	8,16E-06	+
	B4	16,5	4,87E-05	+
	R2	16,3	5,41E-05	+
	T3	23,8	1,07E-06	+

## Présentation synthétique et chronologique des résultats

Nous reprenons dans le graphique suivant le tableau précédent. Nous retrouvons en ligne les 42 candidats pour lesquels des écarts particulièrement élevés (i.e. parmi les cent plus importants) ont été observés (au sens du Chi 2, Cf. Tableau précédent) et en colonne les lots correspondants.

On peut identifier deux groupes distincts de candidats : le premier – de Geneviève Azam à Cécile Verdin – enregistrant des scores anormalement bas dans les lots dépouillés les 14 et 15 juin ; le deuxième – d’Audrey Barral à Jacques Weber – comptabilisant un nombre particulièrement élevé de votes dans les mêmes lots des 14 et 15 juin.

Il est frappant d’observer à quel point les irrégularités sont datées : elles apparaissent massivement dans les lots dépouillés après le 14 juin. Elles apparaissent de plus particulièrement cohérentes : très généralement, lorsque plusieurs irrégularités sont observées pour un même candidat, elles vont toutes dans le même sens lorsque elles sont toutes repérées avant ou après le 14 juin. De part et d’autre de cette date, le sens des irrégularités est généralement opposé.



## Perspectives ultérieures

La perspective retenue jusqu'à maintenant est néanmoins restrictive : on a raisonné marginalement, c'est à dire candidat par candidat. Une analyse plus globale pourrait infirmer les résultats obtenus : si la propension à voter pour le candidat X est positivement corrélé avec la propension à voter pour le candidat Y, il faudrait être plus tolérant dans l'analyse statistique face aux déviations concernant ces candidats si ces déviations vont dans le même sens et être au contraire moins tolérant face à des déviations qui ne vont pas dans le même sens pour ces deux candidats.

Idéalement, le test que nous souhaiterions utiliser devrait tester l'hypothèse suivante : " Est ce que les bulletins du lot L ne diffèrent pas significativement des bulletins des autres lots ? ". Ces bulletins sont constitués de 62 noms de candidats parmi lesquels sont choisi au plus 24

noms. Notons  $\Delta_i = \begin{pmatrix} \delta_i^1 \\ M \\ \delta_i^k \\ M \\ \delta_i^{62} \end{pmatrix}$  avec  $\delta_i^k = 1$  si le bulletin  $i$  correspond à un vote pour le

candidat  $k$ . Si la constitution des lots n'induit pas de tri<sup>1</sup> entre les bulletins, alors les résultats

globaux nous indiquent que  $E(\Delta_i) = \begin{pmatrix} p^1 \\ M \\ p^k \\ M \\ p^{62} \end{pmatrix}$ , où  $p^k$  est le pourcentage de voix obtenus par le

candidat  $k$ . Si on note  $V$  la matrice de variance covariance (de taille 62) de  $\Delta$ , et  $n_L$  le nombre de bulletin du lot  $L$  :

$\left( \frac{1}{\sqrt{n_L}} \sum_{i \in L} \Delta_i - E(\Delta_i) \right)' V^{-1} \left( \frac{1}{\sqrt{n_L}} \sum_{i \in L} \Delta_i - E(\Delta_i) \right)$  tend vers une loi du Chi2 à 62 degrés de liberté si

la constitution des lots n'a pas induit de tri sur les bulletins, et vers l'infini sinon.

Le problème est que la matrice  $V$  est inconnue et ne peut pas être estimée convenablement sur la seule base du fichier des résultats, il faudrait pour cela disposer d'un échantillon de bulletins. L'approche retenue précédemment se contente donc de construire des tests candidats par candidats sur la base de la statistique

$\left( \frac{1}{\sqrt{n_L}} \sum_{i \in L} \delta_i^k - p^k \right) \frac{1}{p^k(1-p^k)} \left( \frac{1}{\sqrt{n_L}} \sum_{i \in L} \delta_i^k - p^k \right)$  qui tend vers un Chi2 à un degré de liberté si la

constitution des lots n'a pas induit de tri sur les bulletins, et vers l'infini sinon.

---

<sup>1</sup> Par cette expression, on entend un tri qui regrouperait des bulletins exprimant les mêmes choix de candidats.

Ces deux approches peuvent être partiellement contradictoires (avec une faible probabilité cependant) : d'un côté on fait un seul test global qui exploite toute l'information sur les comportements de vote des électeurs, de l'autre une grande quantité de test (autant que de cellule "candidat\*lot") ; un test statistique prenant toujours le risque de rejeter à tort l'hypothèse testée (même si les risques pris ici sont de l'ordre de 1/1000 au moins), il peut apparaître quelques sur-représentation dans certaines cellules alors que le test global ne pointerait pas de déviations significatives.

Même s'il nous semble très peu probable que le diagnostique en soit significativement bouleversé (compte tenu notamment des niveaux de risques retenus et de la concentration des sur-représentations dans certains lots), pouvoir disposer d'un échantillon de bulletin permettrait de construire un test statistique plus adapté et produirait une analyse plus sûre.

Une autre piste d'explorations consiste à chercher des facteurs corrélés au profil de certain lots. Un des facteurs potentiel qui vient à l'esprit au vue des premières explorations menées jusqu'ici concerne la date de dépouillement des lots : le degré "d'atypie" des lots dépend-il de la date de dépouillement ? Cette question n'a pas pu être traitée de manière absolument rigoureuse faute de temps, mais elle mériterait une exploration minutieuse. De même, est ce que le degré "d'atypie" est lié à la lettre du lot ou à l'assesseur présent ce jour là. Un degré d'atypie qui serait principalement lié à la date de dépouillement et/ou à l'assesseur présent ce jour là serait particulièrement troublante. Si le degré d'atypie n'est pas significativement lié à la lettre, cela inciterait à penser que ce n'est pas la lettre de l'électeur qui peut constituer un indicateur du vote mais que d'autres facteurs ont contribué à la concentration de certains votes dans certains lots.