

---

# Effets de dates, effets de listes

---

## ANALYSE DES VOTES DES CANDIDATS “MEMBRES ACTIFS” À L’AG D’ATTAC DES 17 ET 18 JUIN 2006

### COMPLÉMENTS

Auteur : Christophe OSSWALD  
Osswald.Christophe@wanadoo.fr

---

## Table des matières

<b>1</b>	<b>Pourquoi une nouvelle étude ?</b>	<b>1</b>
<b>2</b>	<b>Effets de date et indépendance des bulletins</b>	<b>1</b>
2.1	Petites et grandes enveloppes . . . . .	1
2.2	Des Lettres et des jours . . . . .	1
<b>3</b>	<b>Effets de liste</b>	<b>2</b>
3.1	Candidats indépendants? . . . . .	2
3.2	Deux listes, dictées par les candidats et les données . . . . .	2
<b>4</b>	<b>Malhonnête ou malchanceux ?</b>	<b>2</b>
<b>5</b>	<b>Lever les incertitudes</b>	<b>3</b>
5.1	Choix des lettres . . . . .	3
5.2	Choix des listes . . . . .	3

Cette étude fait suite au document publié le 19 juin, et qui a été mis en ligne par Michel Husson :

<http://hussonet.free.fr/attac0.htm>

Ce premier document est essentiellement une mise à l'épreuve des données synthétiques qui ont été fournies lors de l'AG *via* un test du  $\chi^2$ . Il porte sur les segments dits "S1" (bulletins dépouillés le 11 juin) et "S2" (bulletins dépouillés les 14 et 15 juin au matin).

La première conclusion était, que si tous les bulletins arrivés avant le 9 juin au siège avaient été répartis aléatoirement entre les lots S1 et S2, et que la répartition aléatoire des lettres entre S1 et S2 ne faussait que peu cette hypothèse, alors la probabilité d'une telle dissemblance entre les résultats de S1 et de S2 ne dépassait pas une sur vingt milliards – dans le cadre d'une modélisation des effets de liste sur un peu plus de trois listes.

La seconde conclusion était que pour valider, affiner ou invalider la modélisation en listes, il faudrait informatiser les 4960 bulletins de S1 et S2. Le chiffre de "un sur vingt milliards" pouvant varier fortement selon les conclusions de cette étude complémentaire, et les divers choix qui permettaient la définition de S1 et S2.

## 1 Pourquoi une nouvelle étude ?

Il n'y a pas eu de saisie informatique des différents bulletins, mais les résultats "par grande enveloppe" sont disponibles dans le fichier `attac15.xls`. Je tiens à remercier tous ceux qui – indépendamment, et de tous bords – me l'ont fait parvenir. Le fichier joint à ce document, `attac15-chi2.ods` (format `openoffice2.0`), dispose d'un feuillet supplémentaire, qui reprend tous les calculs de  $\chi^2$  présentés par la suite.

Ce fichier apporte quantité d'éclairages sur les résultats intermédiaires. De plus, il invalide largement les sections concernant les "urnes" du premier document. Lors de l'AG, il avait été question de "trois urnes dépouillées avant le 13 juin" (S1) et de "deux urnes dépouillées après le 14 juin" (S2). En fait, il n'y a que des grandes enveloppes, à peu près homogènes en termes de première lettre du nom du votant.

Ces grandes enveloppes, leur constitution, leurs contenus, amènent à considérer différemment nos données.

## 2 Effets de date et indépendance des bulletins

La campagne a été rythmée par divers envois d'information, susceptibles – espérons-le – d'influer sur les votes des adhérents, de telle sorte que le vote "moyen" diffère selon la date d'envoi. Toutefois, le vote moyen sur une période donnée (avant le 9 juin) pour une lettre  $\mu$

ne diffère pas du vote moyen de la lettre  $\gamma$  sur la *même* période.

### 2.1 Petites et grandes enveloppes

Les enveloppes contenant les bulletins sont arrivées au siège, dans l'ordre d'envoi. Elles sont posées dans un grand bac postal. Elles sont réparties, selon le nom inscrit dessus, dans des petits bacs plus petits postaux. Enfin, les enveloppes anonymes sont ouvertes, et remises dans le bac correspondant.

Au vu du fichier `attac15.xls`, une même lettre peut être répartie en plusieurs "grandes enveloppes". Comme à aucun moment de la procédure il n'y a de volonté de mélanger le bac postal de la lettre (ça se ferait entre les étapes (e) et (f) décrites dans le document envoyé par Serge Le Quéau), il n'y a aucune garantie que chaque grande enveloppe, prise isolément, soit représentative de toute la période du vote.

Les grandes enveloppes, ou regroupement de grandes enveloppes, qui peuvent affirmer ou infirmer une hypothèse de fraude sont donc celles qui ne sont pas concernées par ce biais temporel.

### 2.2 Des Lettres et des jours

Nous ne pouvons donc considérer comme indépendantes de la date de vote que les lettres qui sont dépouillées "en une fois".

**11 juin** : A, E, F, H, M, N

**14 ou 15 juin** : B, K, T,

**11 et 14 ou 15 juin** : C, G, J, L, P, R, S, V,

**Trop rare** : I, O, Q, U, W, X, Y, Z, P/M

**Information de répartition perdue** : D,

Seules les lettres A, B, E, F, H, K, M, N et T respectent cette condition d'indépendance entre jour de dépouillement et date d'envoi du bulletin. Seules ces 9 lettres, représentant 1669 bulletins, sont donc prises en considération.

Afin de mener les tests statistiques dans des conditions acceptables, les lettres pesant moins de 50 bulletins ne sont pas prises en compte. Parmi celles-ci, trois sont limitées : O, W et P/M. Ces trois enveloppes sont dépouillées le 14 juin, Nikonoff y est dépassé par 5, 6 et 24 (!) candidats de la liste George. L'enveloppe P/M – dont la signification m'échappe – est archétypale de la liste George.

Ainsi, l'hypothèse de fraude est liée au fait que la liste Nikonoff est significativement mieux placée pour les lettres B, K et T que pour les lettres A, E, F, H, M et N. L'hypothèse de vote honnête est que les différences entre ces lettres ne sont pas significatives.

Le fait que Trouvé ait plus de voix que Nikonoff a souvent été présenté comme un indicateur intéressant.

**Trouvé** : B(14), E(11), F(11), K(14)

**Nikonoff** : A(11), H(11), M(11), N(11), T(14/15)

Nous ne retrouvons pas ici les effets massifs annoncés – et visibles – sur les lots S1 et S2. Ceci accrédite l’hypothèse selon laquelle la séparation d’une même lettre sur plusieurs jours de dépouillement influe sur la répartition des voix à l’intérieur de cette lettre. Toutefois, l’avance de Nikonoff sur la lettre T est considérable, et crée à elle seule un déséquilibre entre le 11 juin et les 14 et 15 juin.

### 3 Effets de liste

Comme expliqué lors du cheminement de l’hypothèse 1 à l’hypothèse 4 de la première analyse, le fait que les différents candidats d’une même liste varient de la même manière est plutôt rassurant qu’inquiétant, et la prise en compte de ce phénomène permet de modérer les inquiétudes quant à l’honnêteté du scrutin.

Je ne présente ici plus que deux hypothèses de listes. La première reprend l’hypothèse 1 : candidats indépendants, pas d’effets de liste. La seconde est proche de l’hypothèse 3 : la liste George, modulée selon les démentis et les candidats “juste cités”, et une liste établie à partir des corrélations avec le score de Nikonoff.

#### 3.1 Candidats indépendants ?

Le test du  $\chi^2$  concerne ici 62 candidats sur 9 lettres. Nous avons donc 488 degrés de liberté. Sous cette hypothèse, le  $\chi^2$  ne devrait dépasser la valeur 563.6 qu’une fois sur cent. Il vaut 781.25, ce qui dépasse très largement ce seuil.

Toutefois, la valeur du  $\chi^2$  est inférieure à ce que nous avons rencontré avec les données de S1 et S2 selon la même hypothèse. Et l’écart avec le niveau admissible est, lui, très inférieur.

#### 3.2 Deux listes, dictées par les candidats et les données

Nous considérons deux listes. La première est directement issue de la lettre de Susan George (colonne AT).

La seconde ne reprend pas la liste des 32 signataires de la contre-lettre (colonne AU) mais les candidats qui sont corrélés positivement à Nikonoff pour les lettres F (forte variation négative) et T (forte variation positive).

A noter que pour la lettre B, Nikonoff fait un très mauvais score (-25 par rapport à la moyenne), mais pour les deux listes identifiées, cela ne traduit rien de cohérent : liste George de -18 à +26, liste Nikonoff de -25 à +25.

**Liste George** : Trouvé, Harribey, Azam, Veilhan, Bénêteau, Baunez, Verdin, Rolet, Mercier, Coulomb, Coiffard-Grosdoy, Blasco, Pradeau, Rouquet, Etchegaray, Cipièrre, Maurin, Clochepin,

Tosti, Jouve, Denimal en entier. Tassi, Silva-Jacinto, Viale, Burier et Dumas sont comptés pour  $\frac{1}{2}$ .

**Liste Nikonoff** : Nikonoff, Tassi, Jonquet, Karbowska, Weber, Ludi, Bernier, Goussot, Lambert, Laurent, Gaziello, Barral, Coignard, Jauffret, Landfried, Gicquel, Maurel, Brustier, Lalot, Faniel, Benoît, Clequin. Par rapport à la liste des 32, on ajoute donc Tassi et Clequin, et on supprime Henry ainsi que les onze candidats les moins bien classés.

On essaie d’expliquer au mieux (minimisation du  $\chi^2$  global) les votes en distinguant des votes par listes au sein des votes individuels. Ces votes par liste estimés sont donnés entre AW68 et BE68 (liste George) et entre AW69 et BE69 (liste Nikonoff). Cela permet d’expliquer l’essentiel des phénomènes sur la lettre T, et une bonne part de ce qui se passe sur F et H.

La lettre B, dont les écarts ne sont manifestement pas liés aux listes, est moins bien expliquée par ce modèle : la saisie des bulletins permettrait peut-être de replacer les écarts constatés sur B dans une liste spécifique, mais équitablement répartie dans les autres enveloppes...

Notons que pour expliquer “au mieux” le vote, les listes ne représentent que 275 bulletins sur 1669 : 16.5%. Le vote aurait-il été très largement un vote par individu plutôt que par liste ? La encore, une analyse des bulletins serait nécessaire pour trancher.

Comme nous rajoutons 18 nombres dans notre tableau de résultats, et que ces nombre optimisent une certaine fonction des données initiales, le nombre de degrés de liberté du système passe de 488 à 470. Le  $\chi^2$  mesuré passe à 527.4 : une probabilité de 3.4%.

Dans cette hypothèse de modélisation des listes, **un scrutin honnête a une chance sur trente de fournir un déséquilibre global de cet ordre.**

Si on s’intéresse à la lettre T de façon isolée, il n’y a qu’une chance sur 2700 d’atteindre le déséquilibre constaté. Fort, mais pas probant. Et si on tient compte du fait que l’on a sciemment pioché l’exemple le plus percutant parmi les neuf exemples disponibles, les chances tombent à une sur 300.

### 4 Malhonnête ou malchanceux ?

Au vu des risques liés à un “faux positif”, c’est-à-dire déclarer truqué un scrutin honnête (encore qu’organisé sans rendre le trucage matériellement impossible, mais c’est un autre débat), la probabilité de commettre un tel “faux positif” d’une chance sur trente doit être considérée comme excessive.

Sur la base des données qui sont indépendantes des effets temporels de vote, il n’apparaît pas statistiquement raisonnable de refuser l’hypothèse selon laquelle

les bulletins dépouillés le 11 juin seraient issus de la même population que les bulletins dépouillés les 14 et 15 juin.

Les distorsions constatées seraient ainsi dues pour partie aux **effets de classement des lettres dans les grandes enveloppes**, et, pour les lettres qui ne sont pas concernées par ce phénomène, à **la simple malchance, dans une mesure raisonnable**.

Afin d'éviter ces déboires dans un scrutin futur, il convient de :

- dépouiller un bac de classement de votants en une seule fois, sans l'étaler sur plusieurs scéances.
- constituer des bacs de classement de façon plus homogènes et plus indépendante des votants : utiliser le dernier chiffre du numéro d'adhérent, ou les deux derniers chiffres.

## 5 Lever les incertitudes

Deux hypothèses fortes ont été faites pour cette analyse : écarter 9 lettres, et constituer deux listes. Si ces hypothèses sont invalidées, il est évident que les conclusions doivent aussi être remises en cause.

### 5.1 Choix des lettres

Si, entre les opérations (*e*) et (*f*) du dépouillement, les enveloppes ont été manuellement et consciencieusement mélangées, alors les grandes enveloppes d'une même lettre, quel que soit la date de dépouillement, devraient avoir la même nature.

Si ces grandes enveloppes ont été constituées au fur et à mesure du tri, lorsque les bacs postaux commençaient à déborder, alors il semble difficile de remettre cette hypothèse en cause, et d'inclure C, D, G,

J, L, P, R, S et V.

### 5.2 Choix des listes

Le liste George est bien définie dans le contexte de la campagne, et explique bien le vote. Il paraît clair que cette liste est un élément précieux et fiable d'explication des bulletins, et qu'elle sera vraisemblablement onfirmée par l'analyse des bulletins eux-mêmes (en tout cas, elle est confirmée par la mémoire, humaine et informelle, de plusieurs des participants au dépouillement).

La liste Nikonoff considérée est choisie pour expliquer les données synthétiques plutôt que le contexte de la campagne. Il faudrait valider cette liste par rapport aux bulletins eux-mêmes. Si elle est notablement différente, il faut s'attendre à ce que le  $\chi^2$  claculé soit plus important.

Il est possible qu'il y ait un ou plusieurs autres phénomènes de liste statistiquement exploitables, qui pourrait concerner la position des candidats sur le bulletin, le sexe des candidats, ou simplement un perception cimmune d'un grand nombre d'adhérents. Cette autre liste éventuelle permettrait peut-être de mieux expliquer les écarts constatés sur les lettres B, H et M, et donc de baisser le  $\chi^2$  calculé.

Il faut malgré tout être conscients que la numérisation des bulletins ne nous amènera pas directement aux listes "naturelles" qu'induisent ces bulletins : le problème sous-jacent est de type NP-difficile, c'est-à-dire que si on est capables de le résoudre efficacement (problème ouvert) alors on sait aussi pirater les cartes bancaires. Mais là, nous quittons les statistiques pour entrer dans les question d'informatique théorique, selon un angle que Gödel n'a pas abordé... mais qui s'applique bien mieux au cas présent que son joli théorème d'incomplétude.

à titre personnel,  
*Christophe Osswald, Attac Brest.*