

T. Coutrot, 2 août 2006

Réponse brève à Michel Lasserre, statisticien débutant

Dans un mail du 31/07, M. Lasserre affirme démontrer que « l'accusation de fraude est injustifiée ». Il présente des calculs qui permettent selon lui de rejeter « l'hypothèse d'homogénéité », qui fonderait les études antérieures concluant à la fraude. Mais M. Lasserre ne sait manifestement pas ce qu'est une loi statistique et ses conclusions sont basées sur des raisonnements techniquement erronés et/ou spécieux.

1. M. Lasserre constate que les (groupes de) lettres IJK, P et R ne donnent pas globalement un résultat plus favorable aux candidats pro Nikonoff que le résultat global ; il en déduit que la présence de lots atypiques au sein de ces lettres ne biaise pas le résultat global (les lots atypiques seraient « compensés par les autres lots de la même série »). (Il reconnaît cependant que les lettres B, L et T sont globalement affectées par leurs lots atypiques).

Le constat empirique de départ est exact, mais la déduction est erronée. M. Lasserre semble croire que l'hypothèse « d'homogénéité des lots » voudrait que tous les lots (et donc toutes les lettres de l'alphabet) donnent exactement les mêmes résultats. C'est ne pas comprendre ce qu'est une loi statistique : l'hypothèse d'homogénéité signifie que dans la mesure où la première lettre de son nom n'influence pas le vote de l'individu, les (inévitables) écarts entre les résultats des différents lots et donc des différentes lettres resteront dans des proportions « normales » (ceci a un sens précis en statistique...), en tenant compte de la taille des lots. Elle n'exclut bien évidemment pas que les résultats présentent des écarts selon les lots et selon les lettres de l'alphabet, mais simplement que ces écarts ne seront pas aberrants.

Dire que le lot P4 (atypique) ne résulte pas d'une fraude parce que la lettre P n'est pas globalement plus favorable que la moyenne à la liste Nikonoff est donc un complet contresens. Le lot P4 est indiscutablement aberrant par rapport aux critères statistiques habituels, et en son absence, la lettre P aurait donné un résultat défavorable (mais pas de façon aberrante...) à la liste Nikonoff au lieu de donner le résultat moyen. N'oublions pas que le basculement du résultat global repose sur le déplacement de seulement quelques points : ainsi les 16 candidats les mieux placés de la liste Nikonoff obtiennent selon le décompte officiel 44% des suffrages alors qu'ils n'en ont plus que 41% si l'on retire les lots atypiques. La fraude a été soigneusement dosée, et elle n'est d'ailleurs repérable que parce qu'elle a porté pour l'essentiel sur un nombre limité de lots (moins d'une quinzaine).

2. M. Lasserre estime que les lettres qui comportent des lots atypiques ne comportent pas plus de « divergences » que les autres. Son critère de « divergence » est un écart de plus de 9 points entre le score d'un candidat dans une lettre et son score global. Par exemple, ce critère l'amène à considérer que la lettre B ne contient aucune divergence, alors que toutes les études antérieures l'identifient comme contenant les lots les plus aberrants (B3 et B4). Mais le choix d'un seuil de 9 points par M. Lasserre est spécieux: les écarts de résultats provoqués par les lots atypiques sont rarement de 9 points, et plus souvent de l'ordre de 4 à 6 points. Si l'on prend le seuil de 4 points au lieu de 9 et qu'on examine le tableau fourni par M. Lasserre, on constate qu'il y a non plus 0 mais 15 écarts au sein de la lettre B, les écarts positifs étant pour la plupart en faveur de candidats de la liste Nikonoff et les écarts négatifs en défaveur des candidats de la liste Susan George... D'ailleurs, plus que le nombre d'écarts à la moyenne, c'est cette systématité de leur orientation pour les lots atypiques qui fonde les conclusions de fraude.

Si j'avais un conseil à donner aux partisans de la thèse de l'accident électoral, ce serait de ne pas perdre de temps à nier l'existence de lots atypiques : n'importe quel statisticien un peu expérimenté les détecte au premier coup d'œil. Ils feraient mieux d'essayer de l'expliquer par des arguments crédibles. Jusqu'à présent, aucune des théories avancées ne tient la route (cf. leur réfutation par Michel Husson , « L'empreinte de la fraude », <http://hussonet.free.fr/empreint.pdf>).

(je ne doute pas que cette réponse circulera sur toutes les listes où le texte de M. Lasserre a circulé)

Incohérences logiques et erreur méthodologique

(réponse de Michel Lasserre à Thomas Coutrot, le 3/08/2006)

Je remercie Thomas Coutrot pour ses précieux conseils à un statisticien débutant, qui m'ont permis d'affiner, de compléter, et de rendre moins contestables mon travail et les conclusions qui en découlent. Cette étude annule et remplace la précédente ("L'accusation de fraude est injustifiée")

Ce travail s'appuie sur un tableau qui se trouve joint à ce mail en fichier pdf.

- chaque colonne concerne l'ensemble des bulletins commençant par la même lettre. Le regroupement des différents lots sous la même lettre présente l'avantage de relativiser certains écarts dus à la petite taille des lots. Pour éviter toute confusion entre les petits lots obtenus et leur regroupement par la première lettre, j'appellerai ces regroupements "lettre". Je précise qu'à l'origine ces "lettres" ne sont pas l'effet d'un regroupement des lots, mais que ce sont les lots qui résultent d'un dégroupement des "lettres".

- le nombre inscrit dans chaque cellule correspond à l'écart entre le pourcentage obtenu dans l'ensemble des bulletins commençant par cette lettre et son pourcentage moyen sur l'ensemble total des bulletins. Cet écart est pondéré en fonction du poids de ce lot par rapport au nombre total de bulletins.

Les 5 dernières lignes indiquent le nombre de candidats qui dépassent différents écarts. Comme le choix de cet écart est essentiel pour déterminer ce qui est acceptable ou non, j'ai compté le nombre de candidats dont l'écart dépassait 0,3, 0,4, 0,5, 0,6 et 0,7.

1) L'incohérence logique des statisticiens "fraudistes"

Le petit tableau suivant montre les "lettres" dont au moins une donnée diverge. J'ai séparé celles comportant des lots prétendues "atypiques" par les experts, et colorées en jaune sur le fichier pdf ("lettres" prétendues atypiques), et celles pour lesquelles aucun lot "atypique" n'a été relevé ("lettres" normales)

Ecart

"Lettres" prétendues atypiques

"Lettres" normales

> 0,3

B, G, IJK, L, PQ, R, TU

C, D, F, H, M, NO, S.

> 0,4

B, G, L, PQ, R, TU

C, H, M, S

> 0,5

B, L, PQ, TU

C, M, S

> 0,6

B

C, M

> 0,7

B

Je ne suis certes pas expert en statistique, mais je suis logicien, et il est logique de penser que la science statistique ne peut que se plier aux règles de la logique. C'est dans ce registre que se tiendra donc mon raisonnement.

La première question qui semble se poser est de savoir quel est le "bon écart". Je la pose donc à l'expert, en lui demandant d'expliquer les raisons de son choix. Mais en réalité peu m'importe sa réponse, elle n'a aucun intérêt pour ce que je vais montrer.

- Si le bon écart était 0,3, on ne pourrait qu'en déduire que les "lettres" prétendues atypiques concernées ne sont ni plus ni moins divergentes que les "lettres" normales. Soit toutes les "lettres" sont atypiques et peuvent toutes avoir été l'objet de fraude, soit aucune ne l'est et aucune présomption de fraude n'est a priori envisageable. On remarquera au passage que V, dont le sous-ensemble V1 est qualifié "d'atypique", ne passe même pas la barre des 0,3.

- On peut bien sûr faire un constat semblable pour les différents écarts jusqu'à 0,6, puisque pour chacun il apparaît des "lettres" "atypiques" et normales. Soit toutes ces "lettres" sont "atypiques", soit aucune.

- Seul 0,7 ne concerne que des "lettres" "atypiques", ou plutôt une seule : B. On pourrait logiquement en déduire que seul un écart de 0,7 permet de déterminer une "lettre" "atypique" qui serait donc sensée contenir des lots "atypiques". Cela signifierait alors que toutes les autres "lettres" ne sont pas suffisamment divergentes pour qu'on puisse les considérer comme "atypiques", et que les lots qui en sont les sous-ensembles ne sont donc à fortiori pas "atypique". Ce qui va à l'opposé des conclusions des statisticiens "fraudistes", et met alors en évidence leur incohérence logique.

2) L'erreur méthodologique de Thomas Coutrot

La démarche de Thomas Coutrot n'est pas innocente, ce qui ne veut pas dire que je remets en cause sa bonne foi car cette démarche n'a certainement pas été guidée par une volonté de tromperie mais par le seul inconscient de son auteur. En s'intéressant en premier lieu aux seuls lots les plus divergents, il (lui ou ses conseillers) a pensé voir un élément pouvant accréditer la thèse d'une fraude. Or, les lots pris individuellement sont des ensembles de bulletins de petite taille, et donc sujets à des divergences bien plus importantes que ne le sont les "lettres". Pour limiter cette "distorsion d'échelle", une approche scientifique privilégiera d'abord l'analyse de l'ensemble le plus grand. Ce n'est que si la "lettre" diverge que l'on peut soupçonner une anomalie et voir si elle provient d'un lot anormal. L'inverse relève non d'un comportement scientifique privilégiant le doute à sa propre conviction, mais d'un comportement partisan cherchant inconsciemment ou non une confirmation de ses désirs au détriment de l'objectivité scientifique. Qu'ils ne le prennent pas mal, mais je me permets de conseiller à Thomas Coutrot et à ceux qui le suivent dans sa démarche de lire "La formation de l'esprit scientifique" de Gaston Bachelard.

3) Incohérence majeure

Il existe seulement trois "lettres" dont l'écart est supérieur à 0,6 : B, C et M.

Ces trois "lettres" sont donc les plus susceptibles d'entraîner une suspicion de fraude. B est pourtant la seule où les "experts" ont relevé des lots "atypiques". On comprend mieux pourquoi quand on s'aperçoit que les écarts de B sont défavorables à la "liste Georges" alors que ceux de C et M lui sont favorables. Si l'on voulait absolument expliquer ces écarts par la fraude, la seule conclusion qu'on pourrait tirer de cette constatation serait que les deux camps auraient simultanément fraudé . . .

Restons sérieux, et admettons alors en toute logique que les écarts observés dans cette élection ne sont pas suffisamment significatifs pour justifier une quelconque fraude.