

En réponse à quelques objections

Michel Husson, 23 juin 2006

NB. les pièces de ce dossier sont en ligne à : <http://hussonet.free.fr/attac0.htm>

Le calcul des probabilités est une discipline difficile. Beaucoup de commentaires sur l'analyse du scrutin oscillent entre deux idées fausses et au fond contradictoires. La première est que tout peut se produire. La seconde est qu'il faut des preuves.

N'importe quoi peut se produire ? Cette idée a été mise en avant par Jacques Testart qui écrivait ceci : « *Je peux seulement contribuer à la dernière polémique par un témoignage sur les tours que nous jouent parfois les statistiques : une étude (évidemment non publiée) avait montré, il y a 15 ans, une relation entre l'ordre alphabétique (la première lettre du nom des patients) et les résultats de leur fécondation in vitro... Cette relation, dite "significative" selon les tests usuels, n'avait évidemment aucune signification et fut contredite dans des séries ultérieures. Mais personne n'a alors prétendu que les résultats de la fécondation avaient été truqués !* »

Cette même idée a été reprise sous une forme voisine par Esther Joly : « *Si le tableau n'était pas aussi triste j'aurais éclaté de rire à la réflexion "anomalie statistique" ou "aberration statistique" comme si lors de dépouillements il n'y avait pas aussi des lois de série...* »

Ces positions ne sont pas scientifiques, et dire cela est tout le contraire d'une affirmation dogmatique. Il ne s'agit pas ici d'invoquer un savoir bétonné, mais au contraire de rappeler la propension au doute comme dimension essentielle de toute démarche scientifique. Le type de configuration que nous avons mis en lumière est typiquement celle qui devrait placer n'importe quel scientifique en position d'alerte. Chaque fois que l'on se trouve en présence de distorsions semblables à celles que l'on a pu observer, on doit en effet se poser des questions et aucun travail scientifique ne tiendrait la route s'il laissait passer sans réagir de telles aberrations.

Il y a dans l'histoire des sciences sociale un exemple célèbre, celui de Cyril Burt. Ce chercheur britannique voulait démontrer que l'intelligence était un caractère purement héréditaire. Il s'était donc constitué un échantillon de jumeaux monozygotes, donc biologiquement semblables, mais qui, pour des raisons liées notamment à la guerre, avaient été élevés séparément et dans des milieux sociaux différents. C'est ce qu'on appelle une « expérience naturelle », d'un grand intérêt en sciences sociales. Après avoir retrouvé ces vrais jumeaux, Burt leur faisait passer des tests de QI qui montraient que les paires de jumeaux obtenaient des résultats voisins, indépendamment du milieu social où ils avaient grandi. Il avait ainsi démontré que l'intelligence était innée et que l'« acquis » comptait pour zéro.

Le problème, c'est qu'il avait bidonné ses résultats. Et la chose intéressante est ici de savoir comment on a pu déceler la supercherie. L'erreur de Sir Cyril (anobli pour la qualité de ses travaux) est qu'il avait dupliqué un certain nombre d'observations, sans changer les résultats de ses test statistiques qu'il avait - bêtement - répété à la décimale près.

Esther Joly n'aurait pas été surprise par cet effet de la « loi des séries », et Jacques Testart, fort de son expérience de l'influence du patronyme sur la fécondation in vitro, n'y aurait sans doute vu qu'un nouvel exemple des coïncidences aberrantes dont il nous a donné un exemple amusant. En effet, il n'est pas impossible de retrouver des résultats strictement identiques, ni de voir se succéder des séries de bulletins copie conforme les uns des autres.

Rien, dès que le hasard intervient n'est strictement impossible. Apprenez à un singe à taper à la machine : la probabilité qu'il rédige à nouveau *A la recherche du temps perdu* n'est pas nulle, donc une telle performance est possible, comme la corrélation significative évoquée par Testart. Cet exemple du singe dactylographe montre que la preuve est impossible à délivrer, et c'est la deuxième idée fautive qui circule sous forme d'un appel à la preuve qui ne peut par définition être délivrée.

Que peut alors la statistique ? Elle peut calculer la probabilité de tel ou tel événement improbable et donner un intervalle de confiance qui va s'exprimer sous la forme d'une proposition formulée par exemple ainsi : il y a une chance sur un milliard que le singe écrive un chef-d'œuvre. Le domaine qui nous intéresse ici concerne l'homogénéité des populations. Prenons l'exemple de sondages électoraux : quand, le même jour ou presque, un sondage donnait le « Non » au referendum à 52% et un autre à 48 %, les statisticiens n'étaient pas étonnés, compte tenu de la fourchette de 4 points qui existe sur un sondage portant sur un millier de personnes. Parce qu'en réalité, le résultat du premier sondage devait être lu ainsi : il y a 95 chances sur 100 que le Non obtienne entre 48 et 56 % selon le premier sondage, et entre 44 et 52 % selon le second, de telle sorte que les deux résultats de 52 % et 48 % se trouvaient au bord de la fourchette, mais dans la fourchette.

En revanche, si un sondage avait donné 52 % au Non, et un autre seulement 36 %, ou au contraire 65 %, tout le monde se serait dit que l'un de ces sondages était mal fait, autrement dit qu'il portait sur un échantillon biaisé. N'importe qui aurait été alerté par cette anomalie, sans pouvoir pour autant apporter de « preuve » de la distorsion. C'est une intuition juste que la théorie statistique permet de calibrer.

Supposons qu'une enquête cherche à mesurer le degré de connaissance des collégiens français, et d'ailleurs de telles enquêtes existent. Supposons maintenant qu'à l'échelle d'un département deux enquêtes successives donnent des résultats très différents. Là encore, on se dirait que c'est bizarre, et qu'on a dû mal définir la population enquêtée. Avec les élections à Attac, nous sommes à peu près dans ce cas de figure : il y a 2 ou 3 paquets de bulletins de vote et ces paquets donnent des résultats très différents.

Or, en raison du classement alphabétique, ces paquets n'ont pas de raison d'être aussi différents. Tel comité, par exemple celui de Brest, a décidé de se concerter sur le vote et a émis probablement des bulletins voisins, la difficulté étant de choisir 24 noms dans une liste où chacun n'en connaît que quelques-uns. Si on avait dépouillé ces bulletins à la suite des uns des autres, on n'aurait pas eu de raison de s'étonner de leur ressemblance. Mais ces bulletins ont été dispatchés par lettres de l'ordre alphabétique, avec dans le cas breton, une probable sur-représentation des « L » (Le Garrec, Le Pen, Le Foll, etc.) qui reste du second ordre.

De la même manière, les bulletins envoyés en dernier sont peut-être différents des premiers, ce qui pourrait induire des effets de série s'ils avaient été classés par ordre d'arrivée. Mais là encore, le tri alphabétique conduit à garantir contre ce type de proximité.

Donc, a priori, les gros paquets de bulletins sont identiques. Ils constituent des échantillons de grosse taille (environ 3000, 2000 et 1000) alors qu'il faut se rappeler qu'un échantillon de 1000 personnes est considéré comme représentatif dans les sondages. Les différences observées entre ces différents paquets constituent une véritable anomalie qui devait légitimement alerter ceux qui ont pu l'observer en premier. Il ne s'agit pas d'un réflexe parano de lecteurs du Da Vinci code qui verraient des complots partout.

Pourtant, on ne répétera jamais assez, la « preuve » statistique parfaite de l'« anomalie » n'existe pas, puisqu'elle correspondrait à une probabilité de 100 % qui est par définition hors d'atteinte des tests statistiques. En revanche, il est possible de calculer

rigoureusement une probabilité et d'établir une proposition de la forme : il y a une chance sur mille que la configuration observée soit « naturelle » et qu'elle s'explique par un pur hasard de la répartition des bulletins. Christophe Osswald a mené ce travail en utilisant la procédure dite « test du Khi-deux » qui permet de répondre à la question : voici deux échantillons, quelle est la probabilité qu'ils soient tirés d'une même population ? Le test s'appuie sur le théorème de Pearson qui établit ceci (je le traduis en langage courant) : soit n résultats de variables aléatoires qui suivent la même loi ; la variable aléatoire qui calcule la « distance » entre les résultats tend à suivre une loi dite du Khi-deux. Ce théorème permet de calibrer les résultats de la manière suivante : si la distance observée dépasse tel seuil (elle est trop grande), alors il y a une chance sur un millier (un million, etc.) que les deux échantillons ne proviennent pas de la même source.

Ce test est donc approprié à notre problème et, appliqué aux résultats dont nous disposons, il montre que la probabilité que les paquets de vote soient tirés d'une même population est très faible. En réalité, les choses sont plus compliquées, car les différents bulletins ne sont pas indépendants ; concrètement, cela veut dire par exemple que les gens qui ont coché Jean-Marie Harribey ont souvent aussi coché Geneviève Azam, et que ceux qui n'ont pas choisi l'un n'ont pas non plus choisi l'autre. Le test du Khi-deux ne s'applique donc pas directement, et il faut neutraliser ce genre d'effets en imaginant des « listes-types » entre lesquelles les votants ont choisi en pratique. Christophe Osswald a traité ce problème, ce qui le conduit à réduire cette fameuse distance, mais elle reste élevée (172). Or, pour 61 degrés de liberté (nombre de candidats - 1) la loi du Khi-deux indique que si les deux échantillons ont un Khi-deux supérieur à 128,5 on a moins d'une chance sur un million de se tromper en affirmant qu'il proviennent de populations différentes. Comme 172 est supérieur à 128,5 nous sommes donc en-dessous de cette probabilité minuscule d'une chance sur un million.

Ce test ne dit pas quelle est le « bon » échantillon, seulement qu'ils sont très significativement différents. Cet argument a déjà été discuté dans notre document « anomalie ou fraude ? » : le point essentiel étant ici la séquence des segments : Segment 1 (favorable aux « georgiens ») puis Segment 2 (favorable aux « non-georgiens ») puis segment 3 - voire segment 4 - de nouveau favorable aux « georgiens ».

Cette discussion répond aussi à l'argumentation de Jasmer Singh quand il écrit : « *ce que dit Jacques Testart est que le moindre écart peut donner lieu à des suspensions. Une personne connaissant la théorie des probabilités n'y attacherait aucune importance que s'il dépasse un certain seuil* ». On a vu que la théorie des probabilités permet justement d'établir que le seuil au-delà duquel les suspicions sont légitimes a été très largement dépassé. Puis Jasmer Singh poursuit en écrivant que ; « *mais même en dehors de ces seuils, bien que ces écarts peuvent paraître significatifs, ils peuvent n'avoir que de très petites valeurs probantes* ». Ce concept de valeur probante n'a pas de sens : il n'existe pas de « preuve » absolue et on ne voit pas ce que peut signifier une preuve relative qui n'aurait qu'une « petite valeur probante ».

Enfin, Jasmer Singh pointe d'autres phénomènes aussi troublants selon lui : « Ainsi, on peut sans doute, adoptant certaines hypothèses plus ou moins raisonnables, calculer la probabilité pour que deux membres du CS, pratiquement inconnus de l'ensemble des militants d'Attac, obtiennent pratiquement le même score que le président d'Attac, connu de tous.. Mais serait-il justifié - pour moi certainement pas - de procéder sur des telles suspicions simplement parce que quelqu'un, se donnant pour probabiliste (nous en avons un certain nombre au CS), trouve que les chances qu'une telle chose arrive a une improbabilité significative ».

Cette question en est une, mais elle ne se place pas au même niveau. Elle pointe une difficulté évidente qui est la suivante : comment choisir 24 noms dans une liste de 62 noms

quand on en connaît seulement une dizaine ? Il y a là un problème de démocratie que les « professions de foi » plus ou moins elliptiques et toujours flatteuses ne peuvent résoudre. Que des candidat(e)s peu connu(e)s aient fait de très bons scores peut s'expliquer par le bouche à oreille, le lobbying, voire par le hasard. Mais il n'y a aucune raison que ces résultats divergent d'un « segment » de bulletins à l'autre, parce que ces bulletins n'ont pas été classés selon des modalités qui pourraient reproduire ces cercles d'opinion. On peut ainsi imaginer que des candidat(e)s soient peu connu(e)s au niveau national mais reconnu(e)s régionalement. Si les bulletins avaient été classés par régions, alors de très grandes différences de résultats d'une région à l'autre auraient pu être observées sans que cela ne pose de problème. Il est par exemple vraisemblable que Jean-Marie Harribey et Geneviève Azam ont obtenu des scores encore meilleurs du côté de Bordeaux ou de Toulouse. Mais, encore une fois, ce genre de biais a été nettoyé par le classement alphabétique et ne peut être invoqué pour faire contrepoids aux irréductibles anomalies de ce scrutin.