

PEDAGOGIE STATISTIQUE DES SONDAGES ELECTORAUX

Benoît Riandey¹

¹ INED, 133 Bd Davout, F75980 Paris cedex 20 et riandey@ined.fr

Résumé.

Les citoyens français ont des raisons d'être troublés lorsqu'à l'occasion des élections présidentielles¹, en 24 heures, deux instituts privés de sondage publient deux estimations d'intentions de vote divergentes tant en niveau qu'en évolution. Les enquêtes sont-elles en cause ? La taille limitée des échantillons ? Des méthodologies divergentes ? Un malentendu sur les réponses attendues de ces enquêtes ? Notre profession des statisticiens a-t-elle expliqué les éléments utiles pour lever ces doutes ? Non et archi-Non.

Le lundi 12 mars, l'IFOP annonce : « *Croisement des courbes ! Le candidat S en tête au premier tour avec 28, 5% des intentions de vote. L'écart au second tour à son détriment en est réduit de 2 points* ».

Le lendemain, mardi 13 mars, la SOFRES annonce : « *Le candidat H confirme son avance avec 30 % des intentions de vote. L'écart au second tour à son avantage en est accru d'un point* ».

Au vu de ces résultats, les citoyens ont des raisons de voir une confusion troublante ! Celle-ci est accrue par l'imprudence statistique de certains médias au détriment des instituts. Le summum a sans doute été atteint par *Paris-Match* à l'égard de l'IFOP : cet institut a mené pendant toute la campagne une enquête en continue (fâcheusement désignée sous le terme de « rolling ») consistant à enquêter chaque jour un peu plus de 300 enquêtés et à produire une estimation moyenne sur les trois derniers jours. Ainsi, avec la même précision que ses concurrents, l'IFOP publie une estimation par jour, représentée par une fort jolie courbe.

On n'aurait qu'à applaudir si *Paris-Match* s'en tenait là sur son site Internet. Par contre, cet hebdomadaire a cru opportun d'y afficher chaque jour l'évolution des intentions de vote par rapport à la veille. Certes, le public n'aura sans doute pas saisi l'énormité de cet affichage tandis que nous décernerons sans hésiter le bonnet d'âne statistique à *Paris-Match* : du jour au lendemain, les échantillons n'évoluent que de 300 entrants et 300 sortants, les vagues intermédiaires communes s'éliminant par différence. La marge d'erreur associée à cette évolution quotidienne est alors si grande que ce calcul en est absurde et ne pas rend pas service à cette belle enquête (cf. annexe). Comme nous le monterons en section 3, le 12 mars a été la date d'un non événement.

Ces non événements sont à l'origine de cette communication. Mais rendons hommage aux heureux efforts pédagogiques conduits par d'autres médias. Nous en donnerons deux exemples. Avant de débattre de la précision des estimations et de l'indicateur de qualité d'un échantillon, nous rappellerons ce qui mathématiquement distingue une estimation statistique d'un dire d'expert politologue. Ensuite nous constaterons que la demande adressée aux

¹ Cette communication ne traite que des élections présidentielles ; les pronostics concernant les législatives s'appuient bien sur une enquête nationale présentant de grandes ressemblances mais donnant lieu à toute une modélisation par circonscription électorale qui ne relève absolument plus du sondage.

sondages électoraux n'est pas tant de fournir des chiffres que de répondre scientifiquement à des questions qualitatives sur l'état du corps électoral, ce que les statisticiens savent faire en testant ces inférences. Nous concluons par un appel aux instituts et aux universitaires pour favoriser cette pédagogie.

1. Estimations statistiques ou dires d'expert

Dans son dialogue avec les instituts, la Commission des sondages désigne par colonne une estimation des intentions de vote pour l'ensemble des candidats à une élection. Il s'agit bien là d'une estimation statistique de ce que répondraient l'ensemble des électeurs si ce questionnaire leur était posé à tous².

Pour être précis, ceci consiste d'abord à choisir un périmètre démographique : le corps électoral des inscrits en France métropolitaine ou en France, ou seulement les électeurs certains d'aller voter ou sûrs de leurs choix ; puis à assurer un juste équilibre de l'échantillon grâce à une variable de pondération de l'échantillon fondée sur des statistiques socio-démographiques et sur le résultat d'un ou plusieurs votes récents. L'équilibrage politique de l'échantillon s'appuie, en effet, sur une période antérieure grâce à la question « Pour qui avez-vous voté à telle élection passée » ? Il n'y a plus qu'à sortir la distribution ainsi pondérée des réponses de ces enquêtés à la question « Si le vote avait lieu demain, pour qui voteriez-vous ? »

Telle est la définition pratique d'une colonne ou de l'estimation statistique des intentions de vote. Modifier ces chiffres ou les produire autrement ne relève plus de la statistique d'enquête et ne permet plus légitimement de faire référence aux marges d'erreurs des statisticiens mesurant la fluctuation d'échantillonnage due à la dimension limitée de l'échantillon, leurs fameux intervalles de confiance.

Certes, les statisticiens ne devraient pas s'offusquer de ce que les politologues et les médias présentent des résultats modifiés, mais à condition qu'ils ne les fassent pas passer pour des résultats d'enquête et qu'ils leur donnent bien le statut de *dires de politologues* et qu'ils ne garantissent pas la même marge d'erreur statistique³.

D'ailleurs la Commission de sondages, dans son rapport du 2 octobre 2012, admet « une certaine marge de manœuvre dans le choix des résultats publiés ». Bien sûr les statisticiens sondeurs font un choix avisé de leurs statistiques de calage, donc de leur pondération comme du périmètre de l'estimation. La Commission, soucieuse à juste titre de la permanence de la méthode⁴, recommande la continuité de la méthode reconnue.

² Nous n'écrivons pas la "vraie" valeur ; ici n'intervient pas d'évaluation d'une vérité -si elle a un sens- de la réponse de l'enquêté à la question sur le vote à venir. Rappelons que l'intervalle de confiance de l'estimation fournie par les statisticiens est une fourchette ne tenant pas compte du biais résiduel d'échantillonnage, de l'effet spécifique du mode de collecte, de l'exactitude ou de la sincérité des réponses.

³ Si on donne un coup de pouce de seulement 1% à une estimation voisine de 50 %, la marge de 3,1 % ne répond plus exactement à la condition de contenir la valeur à estimer dans 19 échantillons sur 20 (ou avec une probabilité de 95 %), mais avec une probabilité de 90,3 %, soit seulement dans 18 échantillons sur 20 (Gerville-Réache, 2012) et tableau 2.2 reproduit en annexe.

⁴ Les statisticiens veillent à l'homogénéité de leurs séries chronologiques en prenant soin en cas de changement de méthode de publier à la date du changement les deux estimations issues des deux méthodes. Cette recommandation, appliquée aux sondages électoraux, serait un progrès méthodologique et de transparence.

Outre le coup de pouce, doivent être reconnus comme dires d'expert les choix de :

- « Faire une estimation candidat par candidat » ;
- Changer de « colonne » d'un candidat à l'autre ;
- « Faire exception pour un candidat dans le choix de la pondération » ;
- Veiller seulement à demeurer dans la pseudo « fourchette » définie par la plus faible et la plus forte estimations de chaque candidat parmi toutes les « colonnes » calculées.

Cette notion de « fourchette », absolument anti-statistique, découverte avec nos collègues statisticiens en 2002, avait créé stupéfaction. Encore aujourd'hui en 2012, considérer qu'un nouveau candidat comme le candidat M, doit bénéficier d'une pondération personnelle dénote de l'ignorance de ce que la pondération concerne l'échantillonnage et non les variables dites d'intérêt, ici les intentions de vote. Pour être plus clair, le redressement de l'échantillon concerne les candidats ou partis du passé pris en référence, pas ceux de l'élection en cours de campagne.

On est donc très surpris de ce que la Commission admette un écart au principe « du même critère de redressement pour tous les candidats testés ». Mieux vaut un vrai coup de pouce assumé publiquement qu'une hérésie statistique.

On comprend bien les doutes de sincérité que suscitent les déclarations de vote pour certains candidats neutres ou refuges et donc le fait que les politologues jugent fondé de déprimer ces intentions au profit de candidats moins communément avouables. Si un modèle de correction stable dans le temps peut être mis au point, on demeure dans l'ordre des modèles et estimations statistiques. Mais si la variabilité conjoncturelle de ces comportements l'emporte, la statistique est impuissante à compenser ces dérives. On se situe alors dans l'expertise et non plus dans la statistique.

La difficulté de ces enquêtes peut conférer l'avantage aux experts, mais il est utile qu'on sache si le propos tenu est parole de sondeur ou parole d'expert et la distinction essentielle entre ces deux rôles n'est, pour des raisons stratégiques, visiblement pas faite dans nos instituts de sondage privés. En tout cas, nous statisticiens, même après la campagne électorale, nous ne savons jamais, sauf confiance, s'il s'agit d'estimations statistiques ou de propos d'expert et nous le déplorons vivement.

2- La précision des estimations statistiques d'intentions de vote

Le *Que sais-je ?* d'Anne-Marie Dussaix et Jean-Marie Grosbras (cf. Tableau 2.2 en annexe issu de leur ouvrage) et l'annexe pédagogique de Pascal Ardilly au rapport du Sénat (Portelli et Sueur, 2010) ont présenté la notion d'intervalle de confiance avec beaucoup de clarté, mais ils ne précisent pas les conditions spécifiques de ces estimations électorales et leurs conséquences en termes de biais et de niveau de précision. Nous devons en préciser le contenu pédagogique.

2.1 Correction du biais de l'échantillon

Pour les sondages électoraux, les estimations reposent sur le recours à une pondération par calage⁵ sur les données sociodémographiques les plus récentes du recensement et sur le/les

⁵ Nous évitons le terme négatif de redressement car ce calage s'appliquerait aussi aux échantillons aléatoires obtenus en l'absence de toute non réponse (situation inconnue dans les pays développés).

Ainsi on éviterait aussi l'expression statistiquement absurde de « redressement des résultats » : on repondère un

dernier(s) vote(s) pertinent(s). Il s'agit donc d'une mesure des transferts d'intentions de vote entre les candidats de ces votes antérieurs et ceux du vote présent ... en tenant compte des entrées et sorties du corps électoral. Notons tout de suite que les seconds tours d'élections, calés sur les résultats du premier tour évitent ce problème crucial de l'évolution de la démographie des électeurs.

Au contraire, dans une élection de premier tour, le calage sur un vote antérieur ne devrait théoriquement porter que sur les enquêtés inscrits à l'élection de référence après l'impossible soustraction du vote -inconnu- des électeurs sortis du fichier électoral. Il s'agit d'une des difficultés incontournables de ces enquêtes, source d'un éventuel biais résiduel d'échantillonnage.

Néanmoins, la disposition de cette statistique officielle des résultats des élections passées sauve la mise des enquêtes d'intentions de vote : en face à face ou au téléphone, apparaît un très fort biais d'échantillonnage (au profit du centre gauche) ; ce biais de sondage très défavorable est compensé par la disposition de cette miraculeuse statistique électorale si liée à l'objet même de l'enquête⁶.

Ce calage conduit à deux effets bénéfiques liés : rectifier la statistique de la distribution partisane lors des élections de référence et réduire les intervalles de confiance des estimations pour la présente enquête, et ceci d'autant plus fortement que la variable estimée est plus liée à la variable de calage⁷. Ceci conduit donc à un arbitrage entre ancienneté et type d'élection ou à un empilement des variables de calage et cet arbitrage peut être inégalement efficace en termes de précision selon les candidats. Mais, en statistique, les variables choisies in fine s'appliquent bien sûr à toutes les modalités de la variable d'intérêt (donc à tous les candidats).

Les seconds tours d'élection présidentielle bénéficient de l'exceptionnelle corrélation entre le vote des deux tours, non seulement en raison de la proximité des deux scrutins, mais du fait de leur objet tout à fait commun : 55 % des votants du premier tour ont voté pour un candidat encore présent au second tour et, pour les autres, les reports sont largement fléchés. Ainsi ai-je calculé que pour la présidentielle 2007, un échantillon de 1000 intentions exprimées conduisait à une estimation d'une précision de 2,3 % au lieu de 3,1 %. Bref il vaut un échantillon classique de taille 2000.

Ces éléments absolument inconnus du débat expliquent la régulière qualité des estimations de second tour qui échappent à toute polémique, contrairement à celles récurrentes relatives au premier tour (Carriou, 2012). Les conditions statistiques en sont profondément facilitées.

Une question subsidiaire importante concerne l'effet des erreurs -volontaires ou non- de déclaration du vote antérieur. Cette erreur ou dissimulation se reproduit le plus souvent aussi sur l'intention de vote actuelle et de ce fait perturbe peu l'estimation après redressement : un

échantillon mais on modifie des résultats.

⁶ Cette situation si favorable n'est pas tout à fait exceptionnelle : elle se retrouve en présence d'un fichier administratif permanent comme les fichiers fiscaux, par exemple pour estimer les revenus ou rentrées fiscales de l'année n+1 à partir de la base de sondage fiscale de l'année n. L'enquête patrimoine de l'INSEE a bénéficié de cette situation optimale.

⁷ Le gain de précision est celui de la stratification a posteriori si l'échantillon déjà socio-démographiquement pondéré est calé seulement sur une élection ou celui de la régression (logistique) pour un calage multiple. Le gain obtenu par stratification a posteriori ne se distingue de celui d'un échantillon stratifié que par l'ajout d'un terme en $1/n^2$, totalement négligeable quand n est la taille de l'échantillon complet.

sympathisant de l'extrême droite masquant à travers la déclaration d'un vote écologiste ses votes passé et prochain surreprésente le vote écologiste passé (et prochain), donc diminue d'autant la pondération donnée à chaque ancien électeur écologiste, ce qui rétablit l'estimation correcte. Par contre, les autres variables de l'enquête seront perturbées par l'introduction d'idées nationalistes autoritaires parmi les enquêtés se déclarant écologistes.

Autrement perturbant est l'effet de masquage d'un vote infidèle antérieur lors d'un retour à sa famille politique. C'est ainsi que, lors de l'élection présidentielle de 1995, le retour au bercail des sympathisants socialistes infidèles aux législatives 1993, repentis et honteux en 1995 (donc dissimulateurs de leur vote de 1993) n'a pu être évalué à son importance et que tous les instituts de sondage ont sous-estimé le vote Jospin, arrivé en tête au premier tour contre le pronostic unanime des instituts de sondage. Ce ne sont pas les « pécheurs repentis » mais les instituts qui ont été pénalisés par ce biais irréductible !

La dissimulation du vote passé est donc traitée comme un déficit ou un excès de sympathisants passés dans l'échantillon. Appliquer cette correction aux nouveaux électeurs revient à faire des hypothèses fragiles de ressemblance de leur comportement avec celui d'anciens électeurs.

Cette sociologie du comportement à l'enquête des électeurs est sujette à une très forte volatilité conjoncturelle et donc très difficilement modélisable par le statisticien. Il appartient à l'expert politologue de le repérer, puis, s'il y a lieu, de donner un coup de pouce quantitatif fondé sur son seul jugement qualitatif. Son doigté n'en fait pas une procédure statistique. Déontologiquement, il ne devrait pas être ignoré des citoyens ni ultérieurement des collègues⁸.

2.2 Affichage des intervalles de confiance

Malgré les progrès accomplis, la précision des estimations reste assez opaque. Certes, l'intervalle de confiance des statisticiens mesure pour les sondages probabilistes la fluctuation d'échantillonnage due au fait de n'avoir pas enquêté toute la population. Cette incertitude demeure pour les enquêtes par quotas et s'y ajoutent les biais de sélection de enquêtés, sans compter celui lié aux déclarations erronées des enquêtés. C'est donc une valeur indicative plutôt minimisée.

Bien sûr certains instituts n'en restent plus au fameux $\pm 3\%$ pour un échantillon de taille 1000, mais détaillent cette précision par pourcentages (conformément à la ligne 1000 du tableau 2.2). Néanmoins c'est déjà cette taille 1000, ce dénominateur, qui pose problème : bien souvent l'effectif efficace n'est plus que de 700, une fois écartés les enquêtés non électeurs, non certains de voter ou n'exprimant pas d'intention. Le demi-intervalle de confiance s'élève alors après de 4 %. Ceci n'apparaît, semble-t-il jamais. Si l'estimation porte sur un sous-échantillon, il n'est plus question de 1000 ou 700 unités concernées mais de bien moins, par exemple de 200 avec une précision de $\pm 7\%$. Avec les progrès de l'informatique et compte tenu du caractère élémentaire de ce calcul, cette information pourrait-être mobilisée par un simple clic sur le site de l'institut de sondage. Qui le premier ?

⁸ Il va sans dire que, compte tenu de l'extrême mobilité entre instituts de sondage sans aucune clause d'exclusion, il n'y a aucun secret commercial de fabrication à protéger eu égard à ces pratiques statistiques. De l'avis de certains responsables d'instituts, ce secret de fabrication serait « très mince ».

2.3 Indicateur de qualité de l'échantillon

Nous autres statisticiens, quoique informés de l'évolution favorable récente, jugeons sévèrement notre ignorance forcée quant à la nature des estimations publiées et à l'absence d'information sur les redressements dans la notice très pauvre accessible auprès de la Commission des sondages. Nous nous étonnons aussi d'avoir appris que la Commission ne demande pas aux instituts de fournir à ses experts l'indicateur évident mesurant la qualité des échantillons. Il s'agit de la distribution de la pondération individuelle des enquêtés et de façon synthétique son coefficient de variation⁹. Un coefficient de variation élevé indique qu'il a beaucoup fallu redresser l'échantillon et des poids dispersés tendent à accroître les fluctuations d'échantillonnage.

Ce coefficient de variation devrait être accessible aux statisticiens sur le site des instituts car il ne troublerait pas le public, en tout cas dans la notice accessible auprès de la Commission.

Si un statisticien d'enquête était présent et actif comme membre de la Commission, il aurait certainement su demander aux instituts cette information de base.

Au contraire, nous rejetons tout à fait l'idée de publier des résultats bruts (c'est à dire non pondérés) car ce sont des chiffres faux qui ne peuvent qu'induire en erreur. C'est bien la colonne de référence qui doit être rendue publique et non ces résultats bruts.

En anticipant un peu sur la suite, notons que, si ces échantillons donnent une vision instantanée globale satisfaisante du rapport de forces entre candidats, ils ont une taille bien insuffisante pour comparer les scores de tous les candidats et encore plus pour mesurer leurs évolutions. La réduction des coûts induite par la collecte par Internet serait l'occasion d'obtenir une précision plus crédible des évolutions en cours de campagne auxquelles s'attachent à juste titre les médias.

Les questions posées à travers les enquêtes ne se ramènent pas à l'estimation des intentions de vote et à sa précision. Il ne s'agit, en effet, pas d'estimations mais d'inférences.

3- Les vraies questions et leurs modes de réponse : estimation ou inférence ?

Les questions que se posent les citoyens ou que traitent les médias sont surtout les suivantes:

- 1- Quelle est aujourd'hui l'intention de vote pour le candidat X ?
- 2- Le candidat X1 serait-il meilleur que le candidat X2 ?
- 3- Le candidat X devance-t-il aujourd'hui le candidat Y ?
- 4- Le candidat X a-t-il progressé depuis la précédente enquête ?
- 5- Les enquêtes A et B fournissent-elles des résultats statistiquement compatibles ?
- 6- Les enquêtes par Internet donnent-elles des estimations différentes des enquêtes téléphoniques ?
- 7- L'enquête finale A fournit-elle des résultats statistiquement compatibles avec le résultat du vote ?

La section précédente répond tout à fait aux questions 1 et 7 en produisant les deux bornes de l'intervalle. La question 2 est relative aux primaires : si le candidat X1 est crédité de 30 % des voix pour l'élection future et son camarade X2 seulement de 28 %, le

⁹ Soit l'écart-type de la pondération divisé par la pondération moyenne.

parti admettra qu'il est plutôt mieux placé et la primaire ou le bureau du parti en décideront sans invoquer l'intervalle de confiance... à moins que X2 ne soit statisticien.

La question 3 teste l'avantage d'un candidat sur ses adversaires. Elle est bien à l'origine de cette communication : le candidat S devance-t-il le candidat H ? Ou, s'il est bien avéré que le candidat H était antérieurement mieux placé que le candidat S, « *Les courbes se sont-elles croisées*¹⁰ » ?

La vraie question n'est évidemment pas la question « *Les courbes de l'IFOP se sont-elles croisées ?* » à laquelle la réponse est évidemment OUI, mais celle-ci « *Les intentions de vote dans la population antérieurement favorables à H, le sont-elles aujourd'hui à S ?* ».

Qu'aujourd'hui les intervalles de confiance des intentions relatives à H et S se recoupent ou non, n'apporte pas toute la réponse cherchée¹¹. Il s'agit donc de vérifier si la différence $\Delta = p_S - p_H$ est *significativement* positive ou non. Un peu de statistique s'impose pour répondre.

Compte tenu de la taille suffisante de l'échantillon, cette variable supposée estimée sans biais peut être considérée comme gaussienne centrée. Sa variance n'est pas triviale, faute d'indépendance entre ces deux pourcentages¹² issus du même échantillon :

$$V_{\Delta} = P_S (1 - p_S)/n + P_H (1 - P_H)/n + 2 P_S P_H/n$$

La table de Gauss indique qu'avec un échantillon de 1000 intentions exprimées, un écart de 3 points au bénéfice de S au premier tour ($p=0,30$) n'était pas du tout extrapolable à la population des électeurs. Il ne permettait pas de conclure sur l'avance de S ni sur le croisement des courbes. Il aurait fallu un écart de 4,8 %, ou pour un écart de 3 points un échantillon de taille 2500.

Dans leur jargon, les statisticiens disent que cet écart n'est pas significatif au seuil de 5 %. Ceci signifie que dans la situation où les deux candidats S et H seraient à égalité, plus de 5% des échantillons donneraient cette fausse avance à S. Un écart non probant !

Lors d'un second tour serré ($p= 0,50$), il faudrait un écart de 6,2 % pour que l'avance d'un candidat soit jugée statistiquement probante (ou significative) à partir d'un échantillon de 1000 intentions exprimées ou pour cet écart de 3 % un échantillon de 4300.

Ces intervalles de confiance rapportés aux courbes présentées par Daniel Boy et Jean Chiche (2012) montrent qu'au cours de la campagne les courbes des candidats H et S ne se sont jamais significativement croisées. Elles se sont seulement statistiquement indifférenciées à certains moments. Pas de quoi faire autant la une des médias, ce qu'interdit le code éthique de la BBC anglaise (Vedel, 2011).

¹⁰ Il va sans dire que si les intentions étaient bien proches et qu'on ne puisse pas dire quelle courbe dominait dans la population, on pourrait éventuellement dire qui domine aujourd'hui mais pas si les courbes se sont croisées.

¹¹ Le raisonnement probabiliste conclut positivement si les intervalles sont disjoints, mais pas négativement si les intervalles se recoupent.

¹² Si un enquêté déclare une intention de vote pour S, il ne peut le faire pour H. On dira qu'une covariance négative accroît la variance de cette différence. D'où le terme supplémentaire ignoré des instituts $2 P_S P_H/n$.

Mais, par ailleurs, si les estimations P_S et P_H sont très proches et qu'on les remplace par leur moyenne dans la formule précédente, cette formule "complexe" est approchée par la formule si simple $V_{\Delta} = 2P/n$. Miracle !

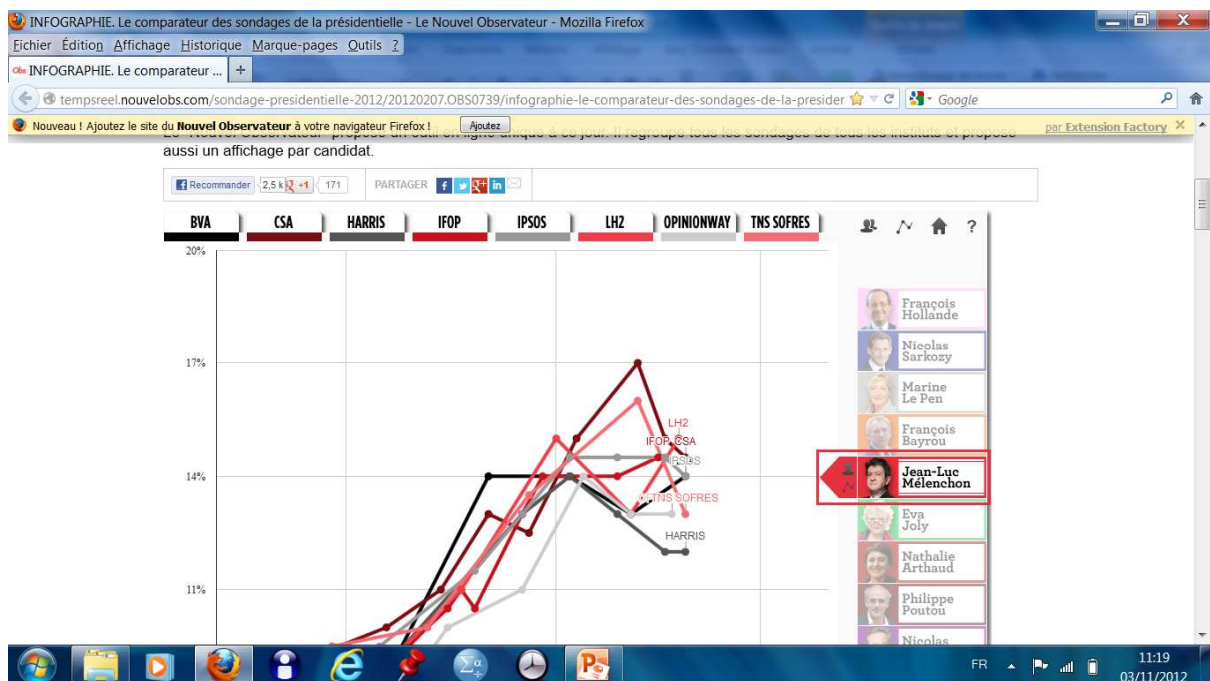
Mais sans parler de croisement des courbes, combien compterait-on d'évolutions significatives au cours de la dernière campagne comme celle du candidat M qui est passée de 8 à 14 % au cours du mois de mars 2012 ?

Les échantillons habituels sont donc bien insuffisants pour différencier les intentions entre candidats au coude à coude et pour suivre à court terme les évolutions d'intentions de vote. Ceci confirme tout à fait le diagnostic exprimé au journal *Le Monde* le 9 mars 2012 par le politologue Jérôme Jaffré "Un sondage isolé ne vaut que s'il est confirmé par d'autres". A moins bien sûr d'avoir directement accru massivement la taille de l'échantillon¹³. Le coût très réduit des enquêtes par Internet en ouvre la perspective, du moins si cette baisse de coût est mise au profit de la qualité.

La recommandation de Jérôme Jaffré a été tout à fait suivie par plusieurs médias qui ont mis en oeuvre des comparateurs de sondage. Ce fut, par exemple, le cas du journal *Le Monde* (sondages IFOP, CSA, OpinionWay, BVA, le 15-16 janvier 2012) et plus largement par le site Internet du *Nouvel observateur* à partir de tous les sondages de huit instituts de novembre 2011 à la veille du second tour, rapprochant ainsi les réponses de près de 10.000 enquêtés (graphique 1).

Graphique 1

Le Comparateur des sondages de la présidentielle



Site du comparateur des sondages :

<http://tempsreel.nouvelobs.com/sondage-presidentielle2012/20120207.OBS0739/infographie-le-comparateur-des-sondages-de-la-presidentielle.html>

¹³ Rappelons que la mesure du taux de chômage à l'enquête Emploi de l'INSEE précise à 0,3 % est menée avec un échantillon (aréolaire) de 100.000 personnes chaque trimestre.

Si on ne peut pas se fier à une enquête isolée, on peut se fier à ce Comparateur de sondages d'une part pour constater la convergence ou non des estimations et des évolutions et surtout pour en faire une statistique de synthèse. La plus précise consisterait à pondérer les 8 enquêtes en fonction de leurs effectifs (ceux collectées par Internet étant plus importants).

Le biais de cette hyper-enquête serait le biais moyen des 8 enquêtes, ce qui est un moindre mal et donc plutôt une garantie. Mais conservons à l'esprit que si les estimations des instituts sont affectées d'un biais commun, comme celui cité pour la présidentielle de 1995, le biais moyen n'est pas réduit par rapport au biais propre à chaque enquête. De même, si le passage à l'isoloir produit un réexamen de l'intention de vote, par exemple en faveur du "vote utile", notre estimation moyenne n'en paraîtra pas moins biaisée que chacune des 8 estimations des instituts.

L'effet propre d'un institut de sondage pouvant être considéré comme constant s'il y a continuité de la méthodologie statistique, les évolutions mesurées par institut, certes imprécises, peuvent être espérées sans biais et l'évolution moyenne entre instituts est alors précise et sans biais.

La question 4 relative à la progression d'un candidat a été posée dès notre introduction. Elle s'avère statistiquement plus simple que de départager la position relative de deux adversaires. L'écart $\Delta = p_t - p_{t-1}$ est calculé à partir de deux sondages successifs, donc habituellement de deux échantillons indépendants, supposés de même taille. Sa variance prend la forme simple :

$$V_{\Delta} = P_t (1 - P_t)/n + P_{t-1} (1 - P_{t-1})/n$$

Sauf évolution massive des intentions, ces deux termes diffèrent peu entre eux. V_{Δ} est alors approchée par l'expression simple $2 P_t (1 - P_t)/n$.

Une évolution de 3 % de la cote d'un candidat sera plus facilement significative qu'une avance de 3 % sur un adversaire en raison de cette formule de variance plus réduite liée à l'indépendance des deux termes. Des échantillons de 1800 intentions exprimées suffiraient au lieu des 2500 utiles pour départager les deux candidats au fameux 12 mars derniers.

Pour en revenir à la bévée de *Paris-Match*, l'échantillon quotidien de 300 unités de l'enquête de l'IFOP affiche une précision de seulement 5,3 %¹⁴. L'évolution depuis la veille pour un candidat de tête de ce premier tour de la présidentielle deviendrait significative au-delà d'un écart considérable de $\pm \sqrt{2} \times 5,3 \%$, soit $\pm 7,5 \%$ ¹⁵.

Au contraire, les variations à 3 jours ou plus reposent sur deux échantillons disjoints complets et indépendants. Leur intervalle de confiance en est réduit d'un facteur $\sqrt{3}$, respectivement $\pm 4,3\%$ et $\pm 6,2\%$.

¹⁴ Se reporter à l'intersection de la colonne 30 % et de la ligne 300 au tableau 2.2 reproduit en annexe.

¹⁵ Les variances, égales, des deux estimations quotidiennes indépendantes s'additionnent, donc doublent. L'intervalle de confiance, fondé sur l'écart-type, racine carrée de la variance, est accru du facteur $\sqrt{2}$, donc de 41%. Mais, pour le second tour disputé de la présidentielle (p=50 %), l'écart devient significatif au delà au delà de $\pm \sqrt{2} \times 5,8 \%$, soit $\pm 8,2 \%$.

Néanmoins la série affichée, bien que sans variation significative, est quotidiennement très stable car alors rapportée à des échantillons de 900 fictifs et non de 300 réels.

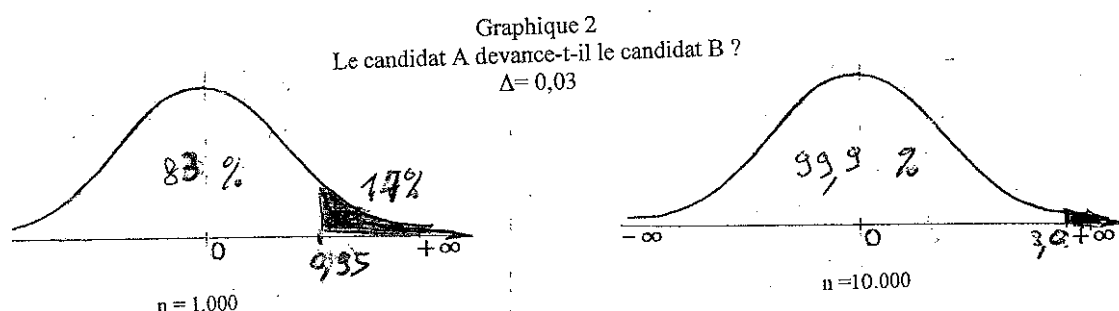
Les questions 5 et 6, traitées par Daniel Boy et Jean Chiche (Boy, 2012) comparent les enquêtes entre elles : les enquêtes d'intentions de vote par Internet ont-elles acquis leurs lettres de noblesse lors de cette élection ? Comme la question 4 abordée en introduction sur l'évolution des intentions pour un candidat, elles portent sur des estimations indépendantes en probabilité car issues d'enquêtes distinctes¹⁶.

4- Quel degré de certitude exiger pour ces inférences ?

Ainsi, sauf à attendre les résultats des confrères ou à réaliser un échantillon de taille économiquement inaccessible, un institut ne pourrait que rarement conclure "scientifiquement" sur la dynamique électorale. Or il serait naïf de penser qu'il faille attendre la confirmation d'une évolution avec une certitude 95 % pour qu'un candidat juge si sa stratégie est positive ou s'il recule. S'il nous fallait ce degré de certitude dans la vie, nous ne prendrions pas beaucoup de décisions. Ce serait dommage car les valeurs centrales de l'intervalle étant les plus probables en l'absence de biais, l'estimation est alors le plus souvent confirmée.

D'ailleurs cette convention des statisticiens de se référer à ce degré de certitude de 95 % est jugée insuffisante par d'autres : les physiciens du CERN ont prolongé de six mois leurs expériences pour renforcer d'une décimale leur certitude de la découverte expérimentale du boson de Higgs. Puis il ont affiché une certitude de 99,999 %.

En définitive, lorsque nous observons un écart de 2% (51-49) ou d'un peu plus de 3 % comme à la dernière présidentielle, quel degré de certitude avons-nous de ce que cette avance soit significative ? C'est en fait LA question à laquelle doivent répondre les sondeurs. Les statisticiens savent très bien le faire. Le graphique 2 l'évoque respectivement pour un échantillon de 1000 intentions exprimées issues d'un institut, puis par l'échantillon de 10.000 synthétisé à partir du Comparateur des sondages du *Nouvel Observateur*. Au vu d'une seule enquête, à la veille du second tour, le candidat H pouvait estimer détenir au moins 83 chances sur cent de précéder son adversaire dans les intentions de vote de la population, mais au vu du Comparateur des sondages, à 99,9 % de chances. Ce n'est pas le boson de Higgs, mais c'est déjà très rassurant !



¹⁶ La concordance ou discordance des estimations relève donc d'un test statistique plus simple, par simple addition des variances, contrairement à la comparaison entre adversaires.

Mais encore une fois relativisons cette certitude et l'excès de confiance dans cette technique du sondage électoral. Ce résultat suppose l'absence de biais d'estimation, d'effet d'isoloir ou de dernier jour. Si ceux-ci se cumulent à 1,5 % au profit du candidat mal placé, le résultat du vote serait inverse à l'attente malgré cette certitude annoncée de 99,9 %.

Chaque institut pourrait livrer aux médias cette estimation de 83 % sur la base de leur seule enquête, plus aisée à transmettre que des intervalles de confiance se recouvrant souvent partiellement et donc non concluants. Indiquer pour chaque fait annoncé (progrès d'un candidat, avantage de l'un, croisement des courbes) son degré de certitude serait compris des téléspectateurs habitués à ce langage par des météorologues confrontés à un hasard beaucoup plus complexe¹⁷. Il reviendrait aux statisticiens politologues d'estimer le gain de certitude -ou le doute- que l'accumulation d'enquêtes apporte dans une situation où il s'agirait bien d'estimations statistiques et indépendantes, donc sans appel téléphonique inter-instituts, situation qu'on espère en vigueur aujourd'hui contrairement à un lointain antérieur.

En conclusion

Nous appelons à davantage de statistique de la part des sondeurs. Pour y parvenir, un effort pédagogique doit être mené par la mise à disposition de fichiers pédagogiques de données individuelles de leurs enquêtes passées. Il est choquant que les étudiants politologues de Sciences-Po n'aient pas accès en salle informatique à de telles données d'enquête.

Un travail scientifique requiert la transparence et la reproductibilité. Les sondages électoraux peuvent-ils s'en prévaloir ? Nous jugeons la transparence insuffisante. L'absence de reproductibilité pour de telles enquêtes d'intérêt public est choquante. A juste titre, Daniel Bachelet déplorait qu'aucun universitaire n'encadre une thèse en statistique fondée sur les données des sondages électoraux. Les instituts auraient-ils le coeur de le leur refuser ?

Bibliographie

- Boy D. et Chiche J. (2012), « L'impact du mode d'administration dans les sondages électoraux », Colloque Sondages2012, session 5.
- Carriou Y. (2012), « sondages électoraux : de la polémique à la pratique », Colloque Sondages2012, session 2
- Chiche J. et Riandey B. « Qualité des enquêtes politiques » (à paraître)
- Commission de sondages « La Commission des sondages face aux élections présidentielles et législatives de 2012 », rapport du 4 octobre 2012.
- Dussaix A.M. et Grosbras J.M. (1993) « Les sondages : principes et méthodes », *Que sais-je ?*
- Gerville-Réache L. (2012). « Sondages d'intentions de vote : l'estimation des « marges d'erreur », Colloque Sondages2012, session 2.
- Portelli H. et Sueur J.P (2010) « Rapport d'information de Sénat n° 54 sur les sondages politiques ».
- Riandey B. (2012). « Sondages électoraux : estimations ou inférences ? ». Journal de l'APMEP.
- Vedel Th. (2012) « L'usage des sondages d'intentions de vote dans le médias », Séminaire de la SFdS du 19 mars 2012 « Le bon usage des sondages électoraux », page Sondages électoraux du groupe Enquêtes de la SFdS (www.sfds.asso.fr)

¹⁷ Cette proposition ne contredit nullement notre appui à la publication pédagogique des intervalles de confiance par assimilation de ces enquêtes à des sondages probabilistes et sous l'hypothèse plus risquée d'absence de biais.

Annexe

Tableau 2.2. – Précision de l'estimation d'une proportion calculée à partir d'un échantillon

Proportion taille observée d'échan- tillon n	5 % ou 95 %	8 % ou 92 %	10 % ou 90 %	15 % ou 85 %	20 % ou 80 %	25 % ou 75 %	30 % ou 70 %	35 % ou 65 %	40 % ou 60 %	50 %
100					8	8,6	9,2	9,6	9,8	10
150				5,7	6,4	6,9	7,3	7,6	7,8	8
200			4,3	5,1	5,7	6,1	6,5	6,8	6,9	7,1
250	2,8	3,4	3,8	4,5	5	5,4	5,8	6	6,2	6,3
300	2,5	3,1	3,5	4,2	4,6	5	5,3	5,6	5,7	5,8
350	2,3	2,9	3,2	3,8	4,2	4,6	4,9	5,1	5,2	5,3
400	2,2	2,7	3	3,6	4	4,3	4,6	4,8	4,9	5
500	2	2,4	2,7	3,2	3,6	3,9	4,1	4,3	4,4	5
600	1,8	2,2	2,4	3	3,3	3,5	3,8	3,9	4	4,1
700	1,7	2,1	2,3	2,7	3	3,3	3,5	3,5	3,7	3,8
800	1,5	1,9	2,1	2,5	2,8	3	3,2	3,3	3,4	3,5
900	1,5	1,8	2	2,4	2,7	2,9	3	3,1	3,2	3,3
1 000	1,4	1,7	1,8	2,3	2,5	2,7	2,9	3	3	3,1
1 500	1,2	1,4	1,5	1,9	2,1	2,3	2,4	2,5	2,6	2,6
2 000	1	1,2	1,3	1,6	1,8	2	2,1	2,2	2,2	2,3
3 000	0,8	1	1,1	1,3	1,4	1,5	1,6	1,7	1,8	1,8
5 000	0,6	0,8	0,8	1	1,1	1,2	1,3	1,4	1,4	1,4
10 000	0,4	0,5	0,6	0,7	0,8	0,9	0,9	1	1	1