

Inférence statistique et sondages électoraux

Benoît Riandey^(*)

Les sondages se sont largement substitués aux jeux de dés comme supports pédagogiques relatifs à la théorie des probabilités. À n'en pas douter les lycéens sont plus souvent en contact avec les communiqués des instituts qu'avec des joueurs de 421. Les sondages qu'ils entendent évoquer sont bien plus souvent les sondages d'opinion et ceux d'intentions de vote que les enquêtes de l'INSEE. Leur mise à l'épreuve par les élections apporte un intérêt pédagogique particulier aux sondages électoraux, malgré les insuffisances théoriques de ces enquêtes. Mais n'est-ce pas la saine confrontation de la théorie idyllique avec la réalité du monde physique ou social ?

L'information sur les sondages électoraux a beaucoup progressé à l'occasion de cette campagne électorale. Les sites des instituts présentent une information relativement riche, mais surtout le Nouvel Observateur a mis en ligne un *Comparateur des sondages de la présidentielle* fournissant depuis novembre 2011 les estimations de tous les candidats pour toutes les enquêtes de tous les instituts. Nous y ferons largement référence.

Le cafouillage médiatique alimenté par ces sondages contradictoires, loin d'écartier les pédagogues de leur examen, constitue un excellent support de questionnement. En tant que statisticien et probabiliste, je suis pour part rassuré par leurs contradictions : avec de si petits échantillons, il serait inquiétant pour leur indépendance qu'ils fournissent précisément les mêmes estimations. Leur convergence soulèverait plus de questions que leurs divergences.

Par contre, les maigres propos tenus par les instituts et plus encore par les médias soulèvent de vraies questions à débattre avec les élèves.

Nous nous proposons de discuter l'évaluation statistique proposée par les instituts, d'énoncer celle qu'un statisticien devrait faire pour un sondage donné ou à l'aide de la collection des sondages opérée par les huit instituts privés actifs sur la présidentielle 2012.

Nous présenterons en annexes les techniques mises en œuvre dans ces enquêtes (annexe 1), les notions de biais (annexe 2) et le cadre réglementaire tendant à réguler cette activité et à produire de la transparence (annexe 3).

Nous montrerons que le débat sondagier ne porte pas sur la valeur et la précision des estimations mais sur les réponses à des questions et sur le degré de certitude de ces réponses. Plus fondamentales que les estimations s'avèrent les inférences. Citons en une série :

(*) Statisticien à l'Institut National d'Études Démographiques et président du groupe Enquêtes de la Société Française de Statistique. Courriel : riandey@ined.fr

- Le candidat potentiel A serait-il mieux placé que le candidat B ?
- Le candidat A progresse-t-il ?
- L'avance du candidat A sur le candidat B est-elle significative ?
- Les deux enquêtes plaçant les intentions pour le candidat A à f_1 et f_2 sont-elles compatibles ? Si oui, avec quel degré de certitude ? Sinon, pourquoi de tels écarts ?
- L'enquête A a-t-elle correctement annoncé le résultat de l'élection ?
- Les enquêtes par Internet donnent-elles des résultats identiques, différents, meilleurs, moins fiables que les enquêtes téléphoniques ?

Les intervalles de confiance fondés sur la loi binomiale et la valeur 1,96 de la loi de Gauss ne suffisent pas pour répondre statistiquement à ces questions. Nous ne sommes pas dans le domaine de l'estimation mais dans celui des tests et l'ensemble de la table de Gauss doit être mobilisée. Si la réponse peut être énoncée simplement, le diagnostic est plus complexe et la justification peu accessible sans mathématiques. On n'en a donc pas trace dans la communication, ni même probablement dans ce travail des instituts français où les statisticiens sont le plus souvent les grands absents.

Discutons de la précision des estimations

Les estimations de second tour s'avèrent précises. Ainsi l'élection du président par 51,65 % des votes exprimés a été excellemment annoncée par les derniers sondages préélectoraux dont les estimations s'étaient entre 52 % et 53,5 %. Si on retient la fourchette de $\pm 3,1$ % correspondant à l'intervalle de confiance au seuil de 5 % pour un échantillon de 1000 enquêtés, tous les instituts ont, en effet, bien capté la vraie valeur dans la marge annoncée.

En fait, six instituts ont enquêté un effectif proche de 1000 électeurs et deux autres, un effectif proche de 2000 conduisant à une fourchette de $\pm 2,2$ %. Les estimations de 52,5 % de ces deux derniers sont plus précises que la moyenne des autres et captent toujours le résultat officiel.

En fait, la fourchette est plus grande qu'annoncée car, sur 1000 enquêtés, seulement de 580 à 680 se sont déclarés certains d'aller voter et ont exprimé leur choix. L'échantillon mobilisé n'est donc que d'environ 600 ou 700 enquêtés. L'intervalle de confiance bondit à $\pm 3,8$ ou 4,1 %. Il est d'autant plus heureux qu'aucune estimation ne se soit écartée de plus de 1,85 % du résultat du vote.

Considérons la profession comme un hyper-institut ayant donc interrogé 10.000 enquêtés. En réunissant leurs efforts dans une estimation commune (et en tenant compte de la taille double de 2 échantillons), l'estimation commune inter-instituts - pondérée par les tailles d'échantillon - vaut 52,75 % pour un intervalle de confiance de ± 1 %⁽¹⁾. Les commentateurs critiques précipités feront remarquer que l'estimation est juste extérieure à l'intervalle [50,65 ; 52,65]. Une situation qui en toute probabilité ne devait intervenir qu'au plus une fois sur 20 élections !

(1) Résultat immédiat en arrondissant le 1,96 de Gauss à 2 et les proportions à 1/2. L'intervalle de confiance vaut alors $\pm 1 / \sqrt{n}$, soit $\pm 0,01$.

Ce serait oublier que l'effectif mobilisé dans l'estimation n'est pas de 10 000 électeurs sûrs de leur vote, mais d'environ 7300 induisant un intervalle de $\pm 1,15\%$, soit [50,5 ; 52,8]. L'intervalle capte de justesse l'estimation commune 52,75 %. L'honneur est sauf !

Néanmoins toutes ces estimations sont supérieures au résultat du vote, ce qui, avec ces huit enquêtes indépendantes, aurait 1 chance sur 256 (2^8) de se produire en l'absence de biais. Anticipant la section 2, nous inférons un biais dont l'interprétation peut relever respectivement d'un biais de sondage ou de couverture, d'un biais de réponse (cf. annexe 2) ou de l'évolution ultime des intentions de vote depuis ces enquêtes généralement collectées le 3 mai. L'observation du statisticien politiste Jean Chiche selon laquelle l'écart entre les candidats semblait se réduire d'un demi point par jour appuie cette dernière interprétation.

Nous nous sommes limités à la seule estimation d'un second tour équilibré, proche de 50 % et donc sujette à l'intervalle maximal de $\pm 3,1\%$ pour 1000 intentions exprimées. Souvent seule référence citée, cette fourchette est symétriquement pessimiste pour des intentions plus fortes comme plus faibles : l'intervalle n'est plus que de $\pm 1,8\%$ pour les candidats ne recueillant que 10 % des suffrages.

Notons enfin que les estimations portent souvent sur des fractions de l'échantillon, par exemple les sous-populations des jeunes, des agriculteurs, des urbains, des Français par acquisition, ... Compte tenu de la petite taille habituelle des échantillons, ces estimations sont très peu précises, sans mention de ce fait par les médias. C'est notamment le cas de l'évaluation des reports de voix des petits candidats au second tour : ainsi, selon les instituts, les estimations de report des voix de J.-L. Mélenchon sur le président élu se sont étalées de 76 à 93 %. La seule mention du fameux $\pm 3,1\%$ est alors fort insuffisante.

Citons encore la présentation médiatique maladroite d'une enquête très intéressante qui a fourni au second tour un excellent résultat, l'enquête en continu de l'IFOP : chaque jour 300 à 350 électeurs sont enquêtés et l'institut établit une estimation portant sur les trois derniers jours, donc sur 1000 enquêtés. Le média client publie quotidiennement l'estimation sur son site, mais y adjoint l'évolution depuis le jour $t - 1$. Or, les deux journées communes d'enquête se neutralisant, l'évolution mesure en fait la différence entre les 300 enquêtés entrant et les 300 sortant, d'une précision chacune au mieux de $\pm 1/\sqrt{300}$, soit $\pm 5,8\%$ (en l'absence de non répondants). Ce média s'évertue là à estimer une évolution quotidienne qui ne sera jamais significativement différente de zéro ! Les critères commerciaux ne sont pas ceux de la statistique.

De la théorie à la pratique : le redressement de l'échantillon

Toutes ces estimations reposent sur la théorie élémentaire des probabilités et des sondages, fondée sur la règle de l'art de Bernoulli, le sondage au hasard mathématique équiprobable et indépendant. Ce modèle est bien éloigné de la pratique des sondages des instituts privés comme en témoigne l'annexe 1.

Prêtons-nous à un propos qui dépasse le programme du lycée : dès la fin de la collecte, l'institut opère deux comparaisons : celle de la structure sociodémographique de l'échantillon à celle du recensement et celle du souvenir du dernier vote au résultat effectif de ce vote antérieur. Le thème très sensible de ces enquêtes conduit habituellement à des écarts très importants entre ces statistiques et leurs estimations par l'enquête. On ne peut donc pas interpréter ce sondage comme un sondage à probabilités égales donnant à tous les citoyens une égale probabilité d'avoir répondu à l'enquête). La théorie des sondages à probabilités inégales prescrit d'accroître en proportion la pondération des enquêtés appartenant aux catégories sous-représentées par le sondage. Cette repondération de l'échantillon est désignée sous le terme maladroit de redressement de l'échantillon. Elle consiste en un simple enchaînement de règles de trois jusqu'à convergence des estimations vers ces statistiques de référence.

Il s'ensuit deux résultats contradictoires. D'une part un sondage à probabilités inutilement inégales⁽²⁾ a tendance à gaspiller de l'échantillon, donc à accroître la variance des estimateurs et donc les intervalles de confiance. Selon l'avis d'experts, cet effet négatif ne devrait pas s'élever à plus de 8 %. La correction du fameux $\pm 3,1$ % ne l'amènerait qu'à $\pm 3,3$ %. Cet effet négatif doit donc être négligé.

Par contre, cette technique de repondération mettant en harmonie une estimation de l'enquête avec une statistique de référence présente un avantage autre qu'esthétique ou de communication : elle améliore la précision des estimations des variables fortement corrélées à cette variable dite de post-stratification. Or, avec le résultat du premier tour de l'élection présidentielle, les sondages électoraux de second tour bénéficient de la disposition exceptionnelle d'une statistique très récente et très corrélée relative à exactement la même population, la liste électorale de l'année. De fait, 55 % des suffrages exprimés au premier tour se sont déjà portés sur les deux candidats du second tour et d'éventuelles consignes de vote ciblent les autres transferts. Cette situation statistique très favorable réduit significativement l'intervalle de confiance, d'environ un tiers. Un échantillon de 1000 enquêtés au second tour approcherait la précision d'un échantillon de 2000 du premier tour. Il semble malheureusement que sur ces questions ne se soit penché aucun universitaire français.

Inférences, avez-vous dit ?

Tout ce propos pour en venir au constat : ce ne sont pas les estimations qui intéressent les citoyens ou les états-majors politiques mais fondamentalement les réponses aux questions « A va-t-il gagner l'élection ? », « A fait-il une bonne campagne ? Progrèsse-t-il ? ».

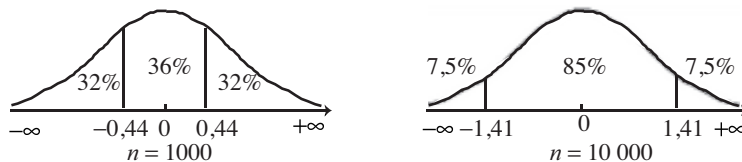
(2) Certains sondages à probabilités inégales peuvent avoir une précision supérieure à des sondages uniformes, notamment pour estimer des variables continues (revenu ou patrimoine moyen). Cette situation est inhabituelle pour estimer des proportions, sauf pour des enquêtes épidémiologiques s'intéressant à des maladies rares inégalement réparties comme le sida.

Le candidat A progresse-t-il ?

La seconde question appelle la réponse la plus simple. Le constat d'une progression du candidat A se teste par la simple différence des intentions en sa faveur à deux enquêtes successives indépendantes. La variance de l'estimation de la différence est alors la somme des deux variances strictement égales. L'intervalle de confiance est donc multiplié par le facteur $\sqrt{2}$, d'où sa formule surprenante $1,96\sqrt{2p(1-p)}/\sqrt{n}$. Le $\pm 3,1\%$ se voit substituer $\pm 4,4\%$.

Il est évidemment bien rare qu'un candidat gagne ou perde $4,4\%$ au cours d'une campagne, ce qui semble enlever toute justification à cette répétition des sondages et devrait amener le plus souvent les médias au commentaire peu vendeur « Question sondages, rien de nouveau ». Toutefois dans la vie courante, nous ne cessons de prendre des décisions moins assurées qu'avec une certitude de 95% . La bonne présentation d'un résultat n'est donc pas d'assortir l'estimation d'un intervalle de confiance au seuil de 5% , mais d'indiquer avec quel degré de certitude l'évolution observée est significative. Revenons au second tour équilibré de la présidentielle et à l'hypothèse d'une progression de Δ du candidat à la traine. Avec quel degré de certitude peut-il croire à sa progression ?

Pour $p = 1/2$, l'écart-type estimé S de cette différence inter-enquêtes de cette estimation $\sqrt{2p(1-p)}/\sqrt{n}$ vaudra $1/\sqrt{2n}$. La variable $\Delta\sqrt{2n}$ est une variable gaussienne centrée réduite. Pour un écart de 1% ($\Delta = 0,01$) obtenu sur un échantillon de 1000 réponses exprimées, elle vaut $0,44$. Selon la table de Gauss, l'estimation issue d'un tel échantillon atteindra ou dépassera cet écart de 1% 32 fois sur 100. Il y a donc deux chances sur trois que ce gain ou cette perte de 1% des intentions de vote soit sans signification (graphique 1).



Graphique 1. Le candidat A progresse-t-il ? $\Delta = 0,01$

Si cet écart résulte de la synthèse des 8 enquêtes totalisant 10 000 réponses exprimées, cet écart ne sera plus atteint ou dépassé que dans 15% des cas (valeurs $\pm 1,41$). Cet écart est significatif 5 fois sur 6. Sans atteindre le degré de certitude scientifique canonique, cette convergence peut alors être prise en comptes par les acteurs politiques⁽³⁾.

Mais la question fondamentale est toutefois de savoir si, dans toute la population des

(3) Si 24% des enquêtés n'expriment pas d'intentions de vote et donc que les estimations ne portent que sur 760 et 7600 unités, $\Delta\sqrt{2n}$ vaut respectivement $0,39$ et $1,23$. Il y a donc respectivement 70 et 22 chances sur 100 que ce gain ou cette perte de 1% des intentions de vote soit sans signification.

électeurs, le candidat A devance le candidat B. Une erreur courante consiste à appliquer la formule précédente. C'est oublier que les proportions P_A et P_B des intentions de vote exprimées en faveur de A et B dans une même enquête ne sont pas indépendantes : si un électeur apporte son suffrage à A, il ne peut aussi l'apporter à B. Il faut tenir compte de la covariance négative entre les deux estimations⁽⁴⁾ :

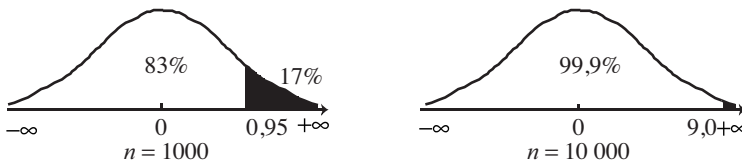
$$V_{\Delta} = P_A (1 - P_A)/n + P_B (1 - P_B)/n + 2 P_A P_B/n.$$

Dans un second tour de l'élection présidentielle, on aurait la relation $P_B = 1 - P_A$ et donc $V_{\Delta} = P_A (1 - P_A)/n$.

Ainsi, par manque d'indépendance des estimations relatives aux deux candidats, l'écart-type et l'intervalle de confiance seraient accrus d'un facteur 2 et non $\sqrt{2}$ comme précédemment. Avec un échantillon de 1000 électeurs, un écart entre les deux candidats ne serait significatif au seuil de 5 % qu'au-delà d'une différence de $\pm 6,2$ %.

Supposons donc qu'une estimation fondée sur 1000 électeurs – anticipant remarquablement le vote – départage les candidats à 51,5 % et 48,5 %. Quelle confiance le candidat en tête aurait-il de l'emporter ?

La variable Δ / S_{Δ} est une variable gaussienne centrée réduite valant 0,95⁽⁵⁾. D'après la table, cette valeur est dépassée seulement dans 17 % des cas. Le candidat en tête a donc 83 chances sur 100 de devancer son adversaire en intentions de vote dans la population ... et de gagner l'élection en absence de biais dans l'échantillon et d'évolution électorale ultime⁽⁶⁾ (graphique 2).



Graphique 2. Le candidat A devance-t-il le candidat B ? $\Delta = 0,03$

Avec l'échantillon cumulé de tous les instituts, soit 10.000 enquêtés⁽⁷⁾, la variable gaussienne Δ/S_{Δ} vaut 3,0 et n'est dépassée qu'avec une probabilité 1/1000. L'avance

(4) Simple application de la formule $V_{X-Y} = V_x + V_y - 2 \text{cov}(X, Y)$

(5) Δ vaut 0,03 ; V_{Δ} vaut $1/n$; S_{Δ} vaut $1/\sqrt{n}$.

(6) Les spécialistes des tests statistiques noteront que ces deux inférences se distinguent bien plus que par l'apparition d'une covariance dans la seconde. La première repose sur un test bilatéral de l'existence d'une évolution : la valeur 0 signifie la probabilité quasi nulle d'une évolution. Dans le second cas, la valeur 0 exprime déjà une probabilité de 50 % de succès pour le candidat (et son adversaire). Il s'agira donc d'un test unilatéral.

(7) Si on tient compte de 24 % d'intentions non exprimées, Δ / S_{Δ} vaut respectivement 0,83 et 2,6. La table de Gauss attribue respectivement à ces valeurs les probabilités 0,2 et 0,004 d'être dépassées. Donc au vu d'une seule de ces enquêtes, le candidat bien placé pourrait croire à sa victoire avec une certitude de 4 sur 5. Au vu des 8 enquêtes, il n'aurait à craindre une défaite qu'avec une probabilité de 4/1000.

du candidat A est assurée avec une probabilité 0,999 ! Si on pouvait faire pleine confiance à la collection des derniers sondages, quelle assurance pour le candidat de tête ... mais les sondages ont surestimé son avance, soit parce que être sondé n'est pas voter ou que l'électeur peut réviser son intention jusqu'à l'isoloir.

En conclusion

Les sondages appellent un commentaire assez professionnel, évitant le va-vite. Ce diagnostic justifie le code de déontologie de la BBC britannique, pays de naissance de la statistique mathématique : Ne pas commencer une émission par un résultat de sondage (*surtout en l'absence de tout fait nouveau statistiquement significatif*).

Annexe 1 La pratique des sondages électoraux

Aujourd'hui les sondages électoraux sont réalisés par téléphone ou par internet. Les sondages téléphoniques ne peuvent plus être menés sans appels de numéros de portables. En l'absence d'annuaire des portables, les instituts génèrent de façon aléatoire des numéros de portables. Le grand nombre d'appels nécessaires à l'obtention d'un questionnaire fait que l'échantillon ne peut être considéré comme aléatoire. Pour encadrer les effets de sélection, les instituts fixent des quotas sociodémographiques à leur échantillon. Ceci signifie que quand l'institut a recueilli suffisamment de femmes ou de jeunes dans l'échantillon, les enquêteurs ne passeront plus de questionnaires dans ces sous-populations.

Pour les enquêtes par internet, on ne sélectionne pas les enquêtés par simple dépôt d'une annonce sur une bannière de peur d'une auto-sélection massive des enquêtés en fonction de leur intérêt pour le thème de l'enquête. Les instituts pré-recrutent des internautes disposés à répondre à des enquêtes de thème non spécifié au recrutement, leur font remplir un questionnaire sociodémographique de recrutement, puis, pour chaque enquête, construisent un échantillon équilibré sur la base de ces informations.

Les enquêtes par internet ont la réputation de fournir des réponses plus sincères par l'absence de dissimulation à un enquêteur.

Annexe 2 Biais de couverture, de sondage ou de réponse

Aucune enquête n'est parfaite. Les enquêtes sont susceptibles d'erreurs tenant soit à l'échantillon, soit à l'information recueillie.

En premier lieu, toute estimation par sondage fait l'objet d'une incertitude d'échantillonnage liée à la taille réduite de l'échantillon et variant à l'inverse de la racine carrée de cette taille. Le respect de la règle de l'art assure d'une estimation sans biais, c'est-à-dire ne tendant pas systématiquement à surestimer ou sous-estimer la quantité ciblée.

Le non respect de cette règle de l'art induit un biais d'échantillonnage qui peut être de deux natures. Soit il s'agit d'un biais de couverture signifiant qu'une sous-population est exclue de l'échantillonnage. Ce peuvent être les enquêtés ne disposant pas du téléphone ou d'internet, les électeurs résidant à l'étranger. C'est d'autre part le fait que des enquêtés ne l'ont pas été avec la même probabilité ou l'ont été avec une probabilité inconnue. Dans ce dernier cas, il aurait fallu pondérer chaque observation par l'inverse de sa probabilité – inconnue – d'inclusion dans l'échantillon. Ces biais d'échantillonnage se corrigent par l'opération de redressement de l'échantillon, procédure opérée sous l'hypothèse qu'à un niveau fin les répondants sont représentatifs des refusants et des personnes impossibles à joindre de mêmes caractéristiques.

Le biais de réponse tient au fait que l'enquêté ne fournit pas la réponse exacte à la question posée. Par exemple, il sous-déclare son revenu, masque son âge ou sa profession, évite de citer un parti politique extrême. La distinction entre biais d'échantillonnage et biais de réponse est d'ailleurs délicate à opérer, par exemple dans l'évocation d'un vote antérieur.

Annexe 3

Cadre réglementaire des sondages électoraux en France

En France depuis 1977, les sondages électoraux font l'objet d'un contrôle visant à assurer que leur sincérité assure le bon déroulement du processus électoral. La loi de 1977 a institué une Commission des sondages constituée essentiellement de magistrats issus des grands corps de l'Etat. Cette commission s'entoure d'experts statisticiens.

Les instituts sont tenus de déposer auprès de la Commission une notice relativement détaillée de la méthodologie du sondage. Si celle-ci n'est pas satisfaisante, la Commission contraint le média concerné à publier une mise au point – obscure – indiquant le manque de fiabilité des estimations.

La révision de la loi en 2002 visait à rendre les sondages plus transparents auprès des citoyens, mais l'information demeure insuffisante. Ce manque de transparence et un abus des sondages par l'Élysée condamné par la Cour des comptes a conduit le Sénat à voter à l'unanimité un projet de loi instaurant davantage de transparence, mais l'Assemblée nationale s'est gardée de l'inscrire à son calendrier de travail. La publication des sondages est interdite le week-end électoral, mais les médias étrangers ne respectent pas cette obligation la législation française.