

« Réguler (enfin) les sondeurs »

par Thomas Piketty*

(Libération, le 8 mars 2011)

Il reste un peu plus d'un an avant le premier tour de l'élection présidentielle. Toutes les semaines, puis tous les jours, nous allons être abreuvés de sondages sur les intentions de vote. Chaque fait et geste des candidats seront interprétés à cette aune. Le problème, c'est que malgré leur sinistre bourde de 2002 (personne n'avait prédit l'éviction de Jospin), les sondages ne sont toujours pas régulés comme ils devraient l'être. En gros, chacun peut continuer de publier les chiffres qu'il veut, sans donner les marges d'erreur, et encore moins les « secrets de fabrication ».

La bonne nouvelle, c'est que si les parlementaires se saisissent du sujet, alors il est temps de muscler la loi sur les sondages et la commission qui va avec. La première modification qui s'impose est d'obliger les instituts de sondage et les médias à publier les intervalles de confiance et non plus un seul chiffre par candidat.

Prenons par exemple le cas typique d'un sondage fondé sur un échantillon initial de 1 000 personnes, avec un taux de réponse de 70 % à la question portant sur les intentions de vote (soit 700 réponses exploitables). Imaginons le cas le plus favorable pour les sondeurs - et malheureusement trop optimiste - d'un échantillon parfaitement représentatif de la population française, sans aucun biais dans les taux de réponse. Un calcul de statistique élémentaire montre que l'intervalle de confiance associé à un tel sondage est de l'ordre de 4%.

Autrement dit, si un candidat A obtient un score de 51% au second tour, alors l'intervalle de confiance est égal à [49 %-53 %]. Cela signifie qu'il existe 95 % de chances pour que le « vrai » score (c'est-à-dire le score obtenu si l'on avait sondé l'ensemble de la population) soit compris dans l'intervalle [49 %-53 %], et 5 % de chances pour qu'il soit en dehors de cet intervalle. Inversement, pour le candidat B obtenant un score de 49 % dans le sondage, l'intervalle de confiance est égal à [47 %-51 %].

Dans la nouvelle loi, les instituts et les médias seraient obligés de publier les deux intervalles [49 %-53 %] et [47 %-51 %] pour les candidats A et B (et non plus les chiffres 51 %, 49 %). Chacun pourra ainsi constater que les intervalles se chevauchent, et donc que le sondage ne nous dit à peu près rien de fiable. Cela va très au-delà de l'habituel reproche selon lequel les sondages ne sont qu'une « photographie » de l'opinion à un instant donné. Le fait est que même si les électeurs ne changent pas d'avis, le sondage n'est pas fiable.

Publier un intervalle de confiance pour chaque candidat aurait un double intérêt. Tout d'abord, on se rendrait compte que la plupart des écarts entre candidats ne sont statistiquement pas significatifs. En particulier, dans le cas du sondage publié le week-end dernier, plaçant Marine Le Pen à 23 %, devant Martine Aubry et Nicolas Sarkozy à égalité à 21 %, il est probable que tous les intervalles de confiance se chevauchent - pour ces trois candidats comme d'ailleurs pour tous les candidats socialistes ! Autrement dit, tout est possible, et il est plus que temps de se concentrer sur les idées et les

programmes plutôt que d'attendre des sondages qu'ils tiennent lieu de débat démocratique.

Le second intérêt d'une telle loi est qu'elle contraindrait enfin les instituts de sondage à rendre publiques les méthodes statistiques détaillées qu'ils utilisent pour redresser les résultats bruts. Les techniques de redressement ne sont pas illégitimes en soi : si l'on observe que les personnes annonçant un vote FN aux sondeurs sont systématiquement deux fois moins nombreuses que les électeurs FN le jour du vote, alors il peut être justifié de multiplier par deux le score FN brut obtenu dans le sondage. Le problème, c'est que ces coefficients de redressement varient énormément dans le temps et suivant les sondeurs... et que les instituts s'obstinent à refuser de rendre publics ce qu'ils considèrent être leurs secrets de fabrications !

Un tel charlatanisme ne peut plus durer. On peut à la rigueur accepter l'argument selon lequel publier les chiffres bruts dans les médias en même temps que les chiffres redressés pourraient être source de confusion, et que l'exposé complet des méthodes de redressement ne peut se faire en quelques lignes. Mais il faut au minimum contraindre les sondeurs à mettre en ligne tous les détails des chiffres bruts et des méthodes statistiques utilisées. C'est d'ailleurs indispensable pour que l'on puisse publiquement vérifier que le calcul des intervalles de confiance prend correctement en compte toutes les marges d'erreur. Et c'est d'autant plus important que les techniques de redressement aboutissent à fortement élargir les intervalles de confiance donnés plus haut...

Pour finir, l'effet d'une telle loi serait d'inciter les sondeurs et les médias à faire moins de sondages, mais à augmenter leur fiabilité. Par exemple, en publiant 5 fois moins de sondages, et en multipliant par 5 la taille des échantillons utilisées (de 1 000 à 5 000 personnes), les marges d'erreur seraient divisées par près de 3. Le débat public y gagnerait en qualité, et les sondeurs en crédibilité.

* Thomas Piketty est directeur d'études à l'EHESS et professeur à l'École d'économie de Paris.