THEORIE DU SONDAGE ALEATOIRE ET ETUDE D'UN SONDAGE D'OPINION AVANT LE 1ER TOUR D'UNE ELECTION PRESIDENTIELLE

Martin KÖRNIG

Résumé

Après le dépouillement d'un échantillon aléatoire, que sait-on sur l'ensemble de la population ? Une solution mathématique à ce problème est proposée sous forme de la densité de probabilité quelle que soit la dimension du problème (nombre de candidats) et quelle que soit la taille du sondage (taille de la population et de l'échantillon). Cette solution amène à une méthode qui permet :

- *une maîtrise des marges d'erreurs statistiques quelle que soit la taille du sondage,*
- *une visualisation des résultats sous la forme des projections,*
- *le calcul de la probabilité pour n'importe quel événement multidimensionnel.*

A partir d'un sondage des intentions de vote avant le 1er tour d'une élection présidentielle en France il convient ainsi de calculer pour chaque candidat :

- *la probabilité d'être élu dès le 1er tour,*
- *la probabilité d'être retenu pour le 2ème tour,*
- *la probabilité d'échec (être ni élu dès le 1er tour ni retenu pour le 2ème tour).*

Mots-clefs : sondage aléatoire, sondage d'opinion, élection présidentielle

1. Introduction

L'échec spectaculaire du candidat Jospin lors du 1er tour de l'élection présidentielle 2002 en France (par ex. Gattegno & Ridet 2002) est l'occasion de se poser la question : Peut-on prédire un tel échec ?

La réponse est « oui » et pour en calculer la probabilité, une solution inhabituelle du problème du sondage est proposée sous forme de la densité de probabilité multidimensionnelle (section 2). Bien que les mots techniques d'un sondage préélectoral soient utilisés pour introduire les paramètres, la théorie développée est bien plus générale.

Etre retenu pour le 2ème tour est un événement multidimensionnel car il s'explique par une relation spécifique des scores de plusieurs candidats et non par le score d'un seul candidat. La stratégie numérique pour calculer la probabilité d'un tel événement est simple :

1. calculer les valeurs de la densité de probabilité dans l'espace multidimensionnel des scores des candidats,
2. stocker ces valeurs avec les scores des candidats correspondants,
3. faire la somme des valeurs représentant l'événement en question.

La définition des événements intéressants dans le cas d'un sondage avant le 1er tour est donnée à la section 3 ainsi qu'un petit exemple numérique (paragraphe 3.4). Dans cet exemple une exploration exhaustive de l'espace multidimensionnel était encore faisable. Pour un sondage de taille plus importante la technique de l'intégration Monte Carlo (par ex. Robert & Casella 1999) semble plus adaptée.

2. Théorie du sondage aléatoire

2.1 La densité de probabilité ρ

Soient :

$\boldsymbol{N} > 1$	la taille du corps électoral, le nombre d'électeurs
$1 \leq \boldsymbol{K} < N$	la taille de l'échantillon, le nombre de personnes interrogées
$\boldsymbol{L} \geq 1$	le nombre de candidats
$0 \leq \boldsymbol{k_i} \leq K$ $(i=1,\ldots,L)$	le nombre d'intentions de vote pour le candidat i présentes dans l'échantillon
$0 \leq \boldsymbol{k_0} = K - \sum_{i=1}^{L} k_i$	le nombre d'abstentions et de votes blancs ou nuls présents dans l'échantillon
$0 \leq \boldsymbol{n_i} \leq N$ $(i=1,\ldots,L)$	le nombre présumé d'intentions de vote pour le candidat i parmi l'ensemble du corps électoral (= nombre présumé de votes pour le candidat i lors de l'élection)
$0 \leq \boldsymbol{n_0} = N - \sum_{i=1}^{L} n_i$	le nombre présumé d'intentions de s'abstenir ou de voter blanc ou nul parmi l'ensemble du corps électoral (= nombre présumé d'abstentions et de votes blancs ou nuls lors de l'élection)

Notre connaissance concernant les intentions de vote de l'ensemble du corps électoral (= prédictions des résultats des candidats lors de l'élection présidentielle) est représentée par la densité de probabilité :

(1a) $\rho : \{0,1,\ldots,N\}^L \rightarrow [0,1[$,

$$
(1b)\quad \rho(n_1,\ldots,n_L) = \begin{cases} 0 & \text{, si } n_0 < k_0 \text{ ou } n_1 < k_1 \ \ldots \text{ ou } n_L < k_L \\ \dfrac{1}{C(N+L,\, K+L)} \displaystyle\prod_{i=0}^{L} C(n_i,k_i) & \text{, sinon} \end{cases}
$$

où

$$
(1c)\quad C(a,b) = \frac{a!}{b!(a-b)!}
$$
(nombre de combinaisons de b éléments parmi a).

Pour le cas particulier $L=1$, on trouve une formule équivalente dans Robert (1992, page 147).

2.2 Expérience aléatoire et ébauche de la preuve

Comme on parle de probabilités, il faut préciser l'expérience aléatoire correspondante. Considérons un ensemble X de N jetons : n_1 jetons de couleur 1, n_2 jetons de couleur 2, …, n_L jetons de couleur L et n_0 jetons d'un autre couleur. Deux ensemble X et X' de N jetons soient considérés différents si $(n_1,n_2,\ldots,n_L) \neq (n_1',n_2',\ldots,n_L')$. Considérons l'ensemble F de tous les ensembles de N jetons qui sont deux à deux différents. Considérons ensuite l'ensemble E composé de toutes les paires ordonnées (X,Y) où $\mathsf{X} \in \mathsf{F}$ et Y est une partie de X de K éléments.

Considérons enfin l'expérience aléatoire de tirer au hasard une paire (X,Y) de l'ensemble E et les événements A et B suivants :

- A : trouver k_1 jetons de couleurs 1, k_2 jetons de couleur 2, …, k_L jetons de couleurs L dans Y

- B : trouver n_1 jetons de couleurs 1, n_2 jetons de couleur 2, …, n_L jetons de couleurs L dans X

La preuve de la formule (1) passe maintenant en 4 étapes. D'abord, le nombre de paires (X,Y) représentant A et B sont :

(2) $\#\,\mathrm{A} = C\,(N+L,K+L)$

et

(3) $\#\,\mathrm{B} = C\,(N,K)$.

Une généralisation de la loi hypergéométrique donne la probabilité conditionnelle de A sachant B :

$$(4) \qquad P(A|B) = \prod_{i=0}^{L} C(n_i\,,\,k_i) \,/\, C\,(N,K)\;.$$

La définition des probabilités conditionnelles P(A|B) et P(B|A) permet enfin d'établir l'identité :

$$(5) \qquad P(B|A) = \frac{P(B)}{P(A)}\,P(A|B) = \frac{\#\,B}{\#\,A}\,P(A|B)\;.\;//$$

Remarques :
1/ Robert (1992, page 147) traite le cas L=1 en tant qu'exemple de la statistique bayésienne. La loi a priori étant évidente dans ce cas, on y trouve une démonstration brève et élégante.
2/ La généralisation de la loi hypergéométrique (4) est mentionnée par exemple par Feller (1968, page 47) qui donne aussi la formule explicite pour L=2.

2.3 Les projections ρ_i

Notre connaissance concernant le candidat i seul (i=1,…,L) ou le nombre d'abstentions et de votes blancs et nuls (i=0) est représentée par la projection :

$$\rho_i : \{0,1,\ldots,N\} \to [0,1[\;\;,$$

$$\rho_i\,(n) = \sum_{\{(n_1,\ldots,n_L)|\; n_i = n\}} \rho(n_1,\ldots,n_L)$$

Les projections ρ_i se prêtent à une visualisation ; les courbes donnent notamment une présentation intuitive des « marges d'erreur » (voir l'exemple paragraphe 3.4).

3. Prédictions avant le 1er tour

3.1 Probabilité d'être élu dès le 1er tour

Le candidat i est élu dès le 1er tour si son score est supérieur à la somme des scores des autres candidats (majorité absolue). La probabilité correspondante est la somme de toutes les valeurs ρ pour lesquelles :

$$n_i > \sum_{j \in L_i} n_j \ ,$$

où

$$L_i = \{1,\ldots,L\} \setminus \{i\} \ .$$

3.2 Probabilité d'être retenu pour le 2ème tour

Pour le candidat i , la probabilité recherchée est la somme de toutes les valeurs ρ pour lesquelles

$$\#\{\, j>0 \mid n_j < n_i \,\} \geq L\text{-}2$$

et

$$\#\{\, j>0 \mid n_j > \sum_{l \in L_j} n_l \,\} = 0 \text{ , où } L_j = \{1,\ldots,L\} \setminus \{j\} \ .$$

La deuxième condition dit qu'aucun candidat soit élu dès le 1er tour.

3.3 Probabilité d'un échec

Une fois les probabilités P_i' (d'être élu dès le 1e tour) et P_i'' (d'être retenu pour le 2ème tour) ainsi calculées, la probabilité d'un échec P_i''' est le complément :

$$P_i''' = 1 - P_i' - P_i'' \ .$$

3.4 Petit exemple numérique

Données : L=4, N=45, K=15

i	k_i
1	0
2	3
3	4
4	5

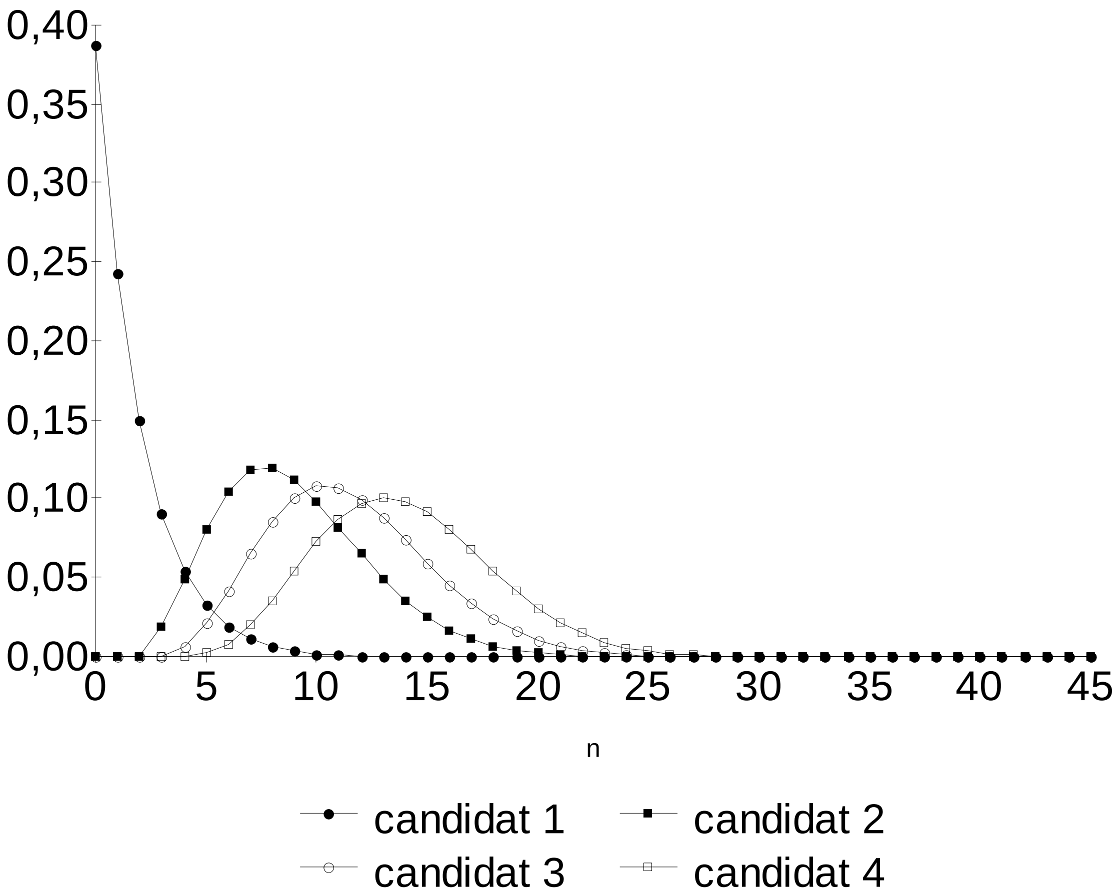
Les quatre candidats rassemblent 12 intentions de vote dans un échantillon de taille K=15 ; il y a donc k_0=3 intentions de s'abstenir ou de voter blanc ou nul. Dans cet exemple, parmi les $(N+1)^L \approx 4500000$ valeurs de la densité ρ, on ne trouve que 46376 valeurs > 0.

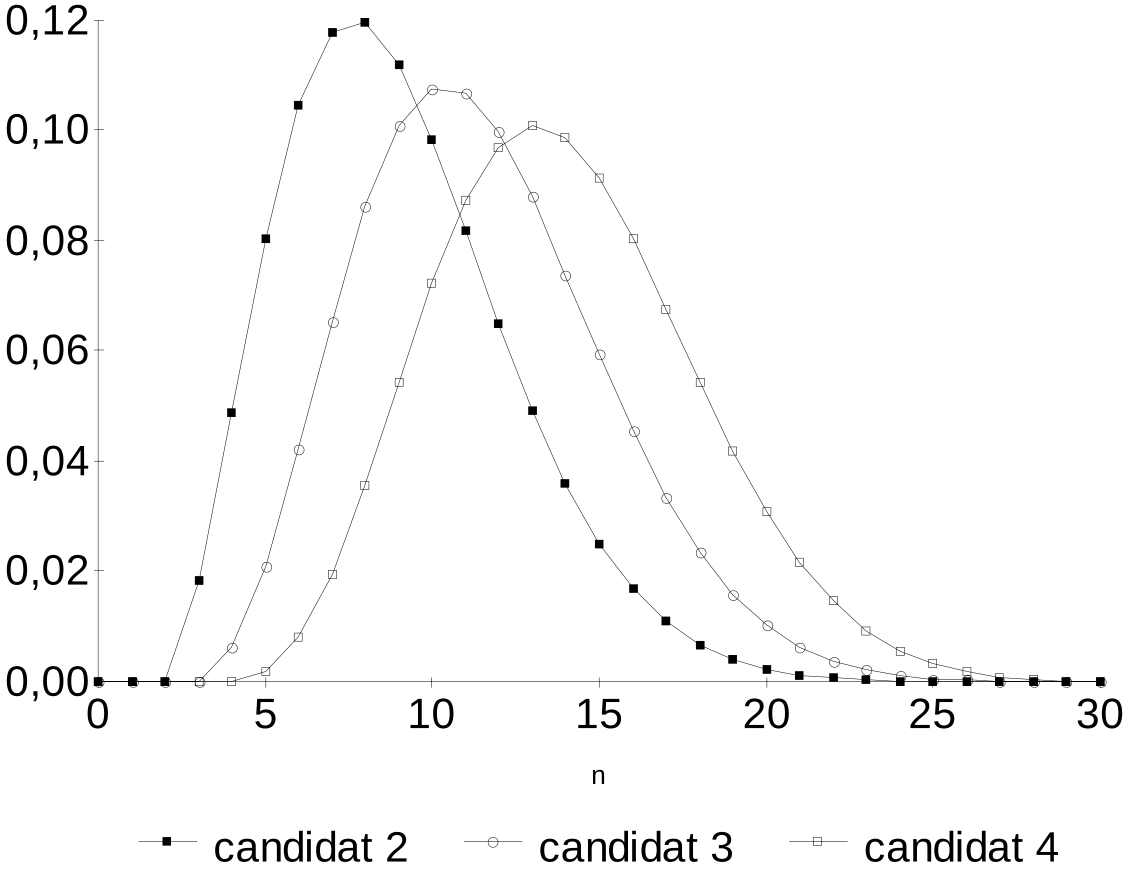
Résultats :

i	Probabilité d'être élu dès le 1er tour	Probabilité d'être retenu pour le 2ème tour	Probabilité d'échec
1	0.0 %	0.5 %	99.5 %
2	0.6 %	31.0 %	68.4 %
3	3.6 %	56.2 %	40.2 %
4	13.0 %	69.7 %	17.3 %

La probabilité d'un 2ème tour est alors 82.8 %.

Visualisations des projections ρ_i :





Références

Feller, W., 1968. *An introduction to probability theory and its applications*. 3rd edition. Wiley, New York.

Gattegno, H. & Ridet, P., 2002. *Le duel Chirac-Le Pen provoque un séisme politique*. Le Monde du 22/04/2002.

Robert, C., 1992. *L'analyse statistique bayésienne*. Economica, Paris.

Robert, C. P. & Casella G., 1999. *Monte Carlo Statistical Methods*. Sprinter, New-York.