

L'espace des phases électoral et les statistiques quantiques. Application à la simulation numérique.

Jean-Louis Rouet, Marc Feix

► **To cite this version:**

Jean-Louis Rouet, Marc Feix. L'espace des phases électoral et les statistiques quantiques. Application à la simulation numérique.. 2005. halshs-00003973

HAL Id: halshs-00003973

<https://halshs.archives-ouvertes.fr/halshs-00003973>

Preprint submitted on 19 Jun 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'espace des phases électoral et les statistiques quantiques.
Application à la simulation numérique.

Marc R. FEIX ¹ et Jean-Louis ROUET ²

ABSTRACT — *We consider different voting scheme giving the collective choice from the lists ranking the preferences of each voter. These different methods do not give always the same result and, consequently, we try to evaluate the frequency of possible conflicts. For that purpose, an electoral phase space of dimension $N = n!$ (where n is the number of candidates) is introduced. Thus all possible lists are taken into account. In this space, we determine the points corresponding to the votes of M electors. Hypothesis on the distribution of these points leads to distributions known, in physics, under the name of Maxwell-Boltzmann and Bose-Einstein. We show how they can be obtained in a numerical simulation with a special attention to the case where N or M or both go to infinity. We end the paper with a study of the frequency of the “Condorcet effect” as a function of the number of candidates and in the limit of a very large number of voters.*

RÉSUMÉ — *On considère des méthodes de vote déterminant le choix collectif à partir des listes de candidats placés par ordre de préférence fournies par chaque électeur. Les différentes méthodes d'agrégation ne donnant pas toujours le même résultat on cherche à déterminer la fréquence des conflits. Pour cela on introduit un espace des phases de dimension $N = n!$ (où n est le nombre de candidats) correspondant à toutes les listes possibles. On détermine ensuite les points représentatifs du vote de M électeurs. Les hypothèses sur la répartition de ces points conduisent à considérer prioritairement les deux distributions connues en physique sous les noms de Maxwell-Boltzmann et Bose-Einstein. On précise alors comment les obtenir dans un problème de simulation en s'intéressant particulièrement au cas asymptotique où M et N tendent vers l'infini. On termine en étudiant la fréquence de l'effet Condorcet en fonction du nombre de candidats dans le cas d'un très large électorat.*

1. INTRODUCTION

La méthode de vote la plus couramment utilisée est le scrutin uninominal majoritaire à un ou deux tours. L'électeur inscrit le nom de son candidat préféré et celui qui récolte le plus grand nombre de voix est élu. Eventuellement un deuxième tour est organisé où seuls subsistent les candidats ayant reçu un nombre suffisant de voix (pour les présidentielles ce sont même les seuls deux candidats arrivés en tête qui sont retenus). Ce second tour permet donc aux électeurs dont le candidat préféré a obtenu un nombre de voix insuffisant d'indiquer un deuxième choix. Il est très probable que ce système favorise une bipolarisation que l'on constate dans beaucoup de démocraties (républicains-démocrates aux USA, travailliste-conservateur en Grande Bretagne, gauche-droite en France . . .) avec parfois, soit des scissions, soit des regroupements dans une des deux composantes.

Toutefois des méthodes plus sophistiquées existent depuis plus de 200 ans initiées par Borda [4] et Condorcet [5]. La mise en œuvre de ces méthodes implique que chaque électeur classe les n candidats par ordre de préférence. Cela donne une masse d'informations très importante. Cependant, le traitement, et même la collecte de cette information, est rendu maintenant possible par le vote électronique. Mais selon la méthode utilisée pour l'agrégation de ces listes on peut, dans certains cas, obtenir des résultats différents et parfois même pas de résultat. Il est intéressant d'évaluer la fréquence de ces difficultés. Pour cela il est nécessaire de se donner un modèle probabiliste donnant les fréquences des différentes listes et d'explicitier les méthodes de simulation. C'est l'objet de cet article dont l'organisation est la suivante.

A la suite de cette introduction, la partie 2 présente l'espace des phases électoral et sa structure (dimension, nature discrétisée ou continue, nombre de points). Partie 3 nous montrons comment cette structure correspond dans les deux cas les plus utilisés aux statistiques quantiques de Maxwell-Boltzmann et Bose-Einstein (et pourquoi celles de Fermi-Dirac ne sont pas à considérer). La partie 4 décrit un modèle d'urne qui donne les trois statistiques (F-D, M-B, B-E). Partie 5 nous donnons quelques résultats portant sur les écarts quadratiques moyens des trois modèles. La partie 6 présente une nouvelle méthode spécialement efficace dans le cas B-E. Les

¹SUBATECH, Ecole des Mines de Nantes, La Chantrerie, 4 rue A. Kastler, B.P. 20722, F-44307 Nantes Cedex 3, France

²Laboratoire de Mathématique, Applications et Physique Mathématique - UMR 6628, Université d'Orléans, UFR des Sciences, F-45067 Orléans Cedex 2, France

parties 7 et 8 s'intéressent aux problèmes soulevés par les cas asymptotiques $M \rightarrow \infty$ et $N \rightarrow \infty$ respectivement pour les cas B-E et M-B. Le réalisme de ces modèles y est également discuté. La partie 9 présente une application portant sur l'évaluation de la fréquence de "l'effet Condorcet" cependant que la partie 10 donne nos conclusions.

2. L'ESPACE DES PHASES ÉLECTORAL

Considérons n candidats. Il est nécessaire de prendre en compte toutes les listes possibles soit $N = n!$. Toutefois dans cette étude nous considérons le cas où N est simplement entier : Il y a deux raisons à cela. La première est qu'une certaine "rationalité" (spontanée ou imposée à l'électeur) pourrait conduire à une réduction du nombre de listes. Par exemple si A est un candidat de "gauche", B du "centre" et C de "droite" on peut espérer qu'un électeur "rationnel" de gauche ne choisit pas la liste ACB ni qu'un électeur de droite la liste CAB . En revanche ABC , BCA , BAC et CBA sont possibles et, ainsi, le nombre de listes passerait de 6 à 4. Cette hypothèse, reliée à la condition de Black [3] est très structurante mais, probablement, assez irréaliste. La seconde raison pour prendre N simplement entier est que, pour fixer nos idées, nous utiliserons les cas $N = 2$, $N = 3$ et $N = 4$ où des représentations géométriques sont possibles et permettent d'obtenir des résultats plus ou moins facilement généralisables. Le deuxième paramètre du problème est M le nombre d'électeurs.

Nous supposons que ces électeurs ont tous le même poids et les mêmes prérogatives (ce n'est pas toujours le cas : au conseil de sécurité de l'O.N.U., par exemple, les membres permanents ont le droit de veto, mais pas les autres). Dans ces conditions, ce qui importe, ce sont les nombres $M_1, M_2, \dots, M_i, \dots, M_N$, d'électeurs ayant choisi respectivement les listes 1, 2, \dots, i, \dots, N . La situation est donc représentée par ces N nombres et l'espace a pour dimension N . En fait, il est plus intéressant de normaliser ces valeurs et de considérer l'ensemble des p_i où $p_i = M_i/M$. Dans ces conditions, le point représentatif se situe sur l'hyper-triangle délimité par l'hyper-plan $\sum_i p_i = 1$ et par les hyper-plans $p_i = 0$ (dans le cas $N = 3$, ce sont des plans et un triangle que l'on considère). Les figures 1 donnent les cas $N = 2$ (fig. 1a), pour lequel le point est sur le segment AB , $N = 3$ (fig. 1b) où le point est sur le triangle équilatéral ABC et $N = 4$ (fig. 1c) où le point est à l'intérieur du tétraèdre dont les 4 faces sont les 4 triangles équilatéraux. Une propriété intéressante apparaît. Considérons le cas $N = 3$ (fig. 1), la probabilité p_3 de la 3^{ème} liste est donnée par Mm . Or, le segment MN , obtenu en abaissant la perpendiculaire de M sur AB , est telle que $MN = \sqrt{3/2}Mm$. Un simple changement d'échelle nous permet donc de rester dans le plan dans le cas $N = 3$ et dans l'espace pour $N = 4$. Dans ce dernier cas, le point représentatif est situé dans le tétraèdre (les 4 hauteurs valant chacune l'unité) et les 4 probabilités sont données par la longueur des segments joignant le point à ses projections sur les 4 faces. Automatiquement on a $\sum_{i=1}^4 p_i = 1$.

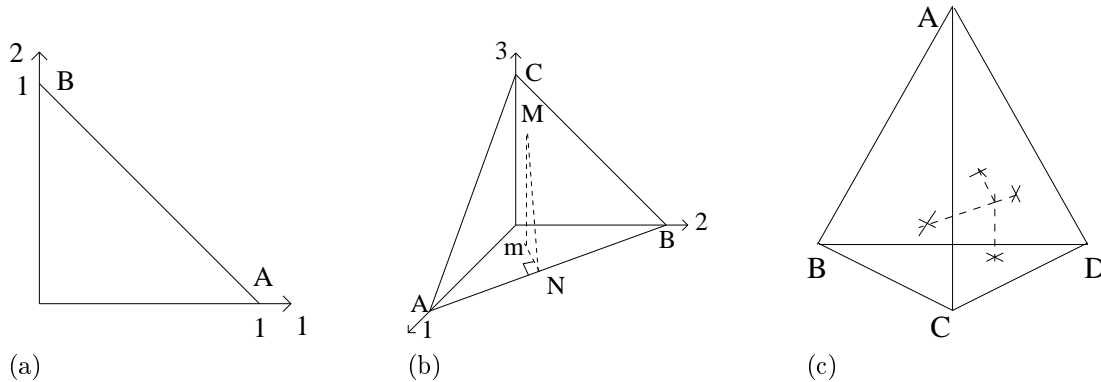


Figure 1: Espace des phases électoral pour un nombre N de listes égale à : (a) $N = 2$, (b) $N = 3$ et (c) $N = 4$.

La façon dont est rempli cet espace de dimension N dépend de N et du nombre M d'électeurs. Si M est fini, seuls un nombre fini de points sont accessibles (cf partie 3). Pour $M \rightarrow \infty$, l'espace des phases est continu et tout point du segment, triangle, tétraèdre, hyper-triangle peut-être obtenu.

3. CAS OÙ M EST FINI

1111	1123	1222	1244	2222	2244	3333
1112	1124	1223	1333	2223	2333	3334
1113	1133	1224	1334	2224	2334	3344
1114	1134	1233	1344	2233	2344	3444
1122	1144	1234	1444	2234	2444	4444

Table 1: Points de l'espace des phases à $N = 4$ dimensions dans le cas de $M = 4$ électeurs.

Que devient cet espace des phases quand M est fini ? Pour le voir traitons tout d'abord le cas $N = M = 3$. Il y a $3^3 = 27$ possibilités mais certaines donnent le même point de l'espace des phases. Par exemple le point central (1 vote pour chacune des trois listes) est obtenu 6 fois (toutes les permutations associant les trois électeurs aux trois listes). Un moyen algébrique d'obtenir tous les points ainsi que leur poids respectif dans la statistique M-B est donné par une généralisation de la formule du binôme qui permet, par exemple, de déterminer la fréquence de pile et face pour M jets. Ici il suffira de développer de préférence avec un langage manipulant l'algèbre.

$$(x_1 + x_2 + \dots + x_N)^M \quad (1)$$

Traitons le cas $N = M = 3$ après avoir rebaptisé x_1 en a , x_2 en b et x_3 en c

$$(a + b + c)^3 = a^3 + b^3 + c^3 + 3(a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2) + 6abc \quad (2)$$

Les différents monômes indiquent les points du triangle : b^3 par exemple indique trois votes pour B ; bc^2 1 vote pour B et 2 pour C ; abc 1 vote pour chaque etc ... Les figures 2 et 3 indiquent la position des points dans le triangle pour $N = M = 3$ et $N = 3, M = 4$: il y a respectivement 10 et 15 points.

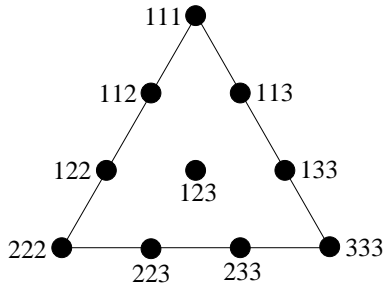


Figure 2: Localisation des points dans l'espace des phases électoral pour $N = M = 3$ (dans le triangle ABC de la figure 1(b)).

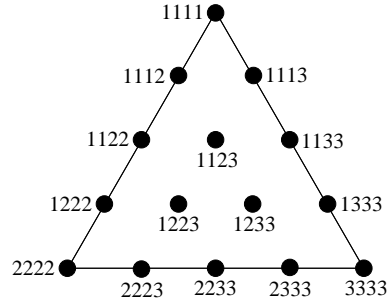


Figure 3: Idem que la figure 2 pour $N = 3$ et $M = 4$.

Un autre moyen de lister les points de l'hyper-triangle est d'énumérer tous les nombres de M digits, chaque digit prenant les valeurs de 1 à N . Toutefois dans notre énumération nous veillerons à ce que les digits successifs n'aillent jamais en décroissant : ainsi, dans le cas $N = M = 4$ après 1124, nous inscrivons 1133. En effet, 1131 qui, à priori, devrait succéder à 1124 a déjà été pris en compte par 1113 à une permutation près des votes des électeurs 3 et 4. La table (1) donne, à titre d'exemple, les 35 points du tétraèdre correspondant à $N = 4, M = 4$.

Rappelons que $ijkl$ est associé à un vote pour les listes i, j, k et l respectivement. La figure 4 représente les différents étages de tétraèdre : la base (0 vote pour la liste 4), le 1^{er} étage (1 vote pour 4), les 2^{ème}, 3^{ème} et 4^{ème} étages, avec respectivement 2, 3 et 4 votes pour la liste 4.

Comment obtenir le nombre de points pour N et M quelconques ? Pour cela on dispose les M électeurs et on place les $N - 1$ séparations qui indiquent le nombre d'électeurs sélectionnant respectivement les listes 1, 2, ..., N (voir figure 5). Sur cette dernière on voit que les listes 1, 2, 3, 4, 5, 6 sont choisies respectivement par 2, 1, 0, 3, 2, 0 électeurs. Combien de configurations obtient-on ? Les $N - 1$ séparations et les M électeurs donnent lieu à $(N + M - 1)!$ permutations mais il faut diviser par $(N - 1)!$ puis par $M!$ pour tenir compte que

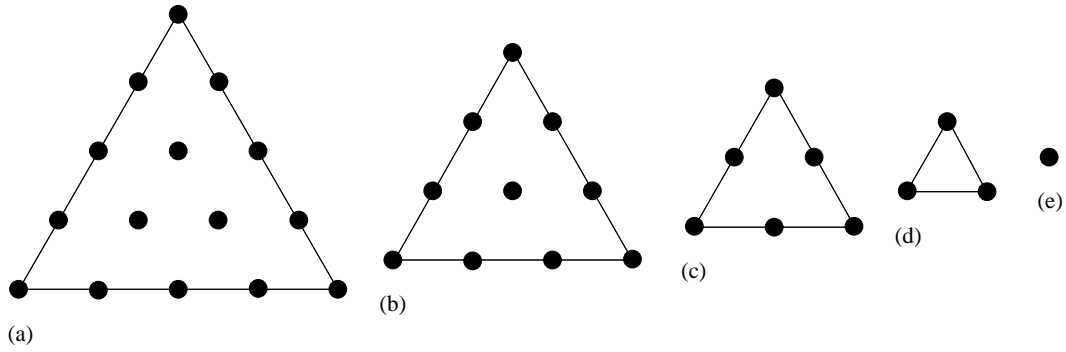


Figure 4: Différents étage du tétraèdre avec $N = M = 4$ (cf figure 1(c)). (a) “Rez-de-chaussée”, (b) 1^{er} étage, (c) 2^{ème} étage, (d) 3^{ème} étage, (e) 4^{ème} étage (1 point).

dans le comptage des points de l’espace des phases toute permutation des électeurs et des séparations donne le même point. Le nombre de points est donc donné par :

$$\frac{(N + M - 1)!}{(N - 1)!M!} \quad (3)$$

On vérifie, bien sûr, que (3) redonne bien le nombre de points obtenus dans les exemples précédents.

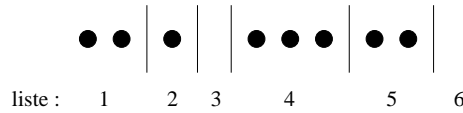


Figure 5: Décomptage du nombre de points de l’espace des phases pour $N = 6$, $M = 8$. Les ronds représentent les M électeurs choisissant l’une des listes représentée par un intervalle délimité par un trait.

Tout physicien a reconnu en (3) une des célèbres formules de la statistique quantique (voir par exemple [8]). De façon plus générale, il est possible d’établir un pont entre la physique statistique et la théorie des votes. Cela commence par le langage et le petit dictionnaire suivant traduit les termes employés par les physiciens en langage “électoral” :

particule	→ électeur
occuper	→ sélectionner
niveau, site	→ liste
stat. Maxwell-Boltzmann	→ “culture identifiable”
stat. Bose-Einstein	→ “culture anonyme”

On voit tout de suite pourquoi les statistiques de Fermi-Dirac ne sont pas considérées en théorie des votes : La phrase définissant cette statistique via le principe d’exclusion dit que *deux particules ne peuvent occuper le même niveau d’énergie*, soit, *deux électeurs ne peuvent choisir la même liste*. Il n’y a, bien sûr aucune raison à une telle interdiction. Quant aux cultures “identifiable” et “anonyme”, il s’agit d’un jargon dont l’origine est assez obscure.

Deux remarques s’imposent :

- Pour N fixé et $M \rightarrow \infty$ les points tendent à remplir uniformément le triangle ($N = 3$), le tétraèdre ($N = 4$) et l’hyper-triangle.
- La statistique de Bose-Einstein consiste à supposer une égale probabilité pour tous ces points.

Il existe, bien entendu, une autre statistique (appelée statistique de Maxwell-Boltzmann). Elle consiste pour des valeurs données de N et M à introduire les mêmes points que pour la statistique de B-E mais au lieu de leur donner à tous le même poids, de supposer des probabilités différentes. En fait, ces probabilités sont proportionnelles aux coefficients des différents monômes obtenus en explicitant (1). Par exemple les cas $N = 3, M = 3$ et $N = 3, M = 4$ redonnent les deux schémas des figures 2 et 3 avec les probabilités indiquées sur les figures 6 et 7.

Reste à traiter le cas de F-D pour lequel il est nécessaire que $N \geq M$. Le nombre de points est tout simplement le nombre de combinaisons possibles de M éléments choisis dans un ensemble en contenant N , soit $C_N^M = N!/[M!(N - M)!]$. Ces points forment donc un sous ensemble de points des statistiques de M-B et B-E et ils sont affectés du même poids (tout comme pour B-E). Pour $N = M$ il n'existe qu'un point $p_1 = p_2 = \dots = p_N = 1/N$. Rappelons que F-D ne joue aucun rôle dans les problèmes de vote.

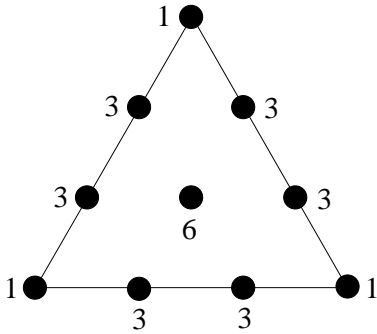


Figure 6: Maxwell-Boltzmann avec $N = M = 3$. Les probabilités ont été multipliées par $27 = 3^3$.

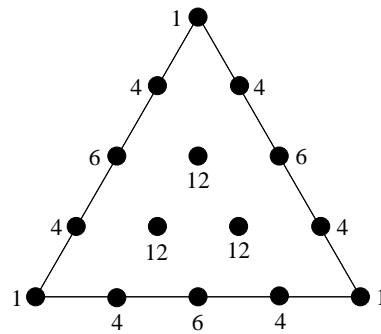


Figure 7: Maxwell-Boltzmann avec $N = 3, M = 4$. Les probabilités ont été multipliées par $81 = 3^4$.

Il est curieux de constater que les deux principales distributions utilisées pour les études probabilistes des votes sont connues sous les noms respectifs de “cultures anonymes” (pour B-E) et “identifiable” (pour M-B). Afin de mieux comprendre la signification de ces deux statistiques, il est intéressant de les connecter à un problème de tirage dans une urne. Nous reviendrons aussi sur leur réalisme partie 8.

4. OBTENTION DE 3 STATISTIQUES PAR TIRAGE DE BOULES DANS UNE URNE

Le cas de M-B est le plus simple à réaliser. On place dans une urne N boules numérotées de 1 à N , chacune correspondant à une liste. Le premier électeur tire une boule, choisit la liste correspondant au numéro de la boule et remplace cette dernière dans l’urne, les autres électeurs font de même. Il est alors clair que si $M \rightarrow \infty$ on se rapproche arbitrairement près du point $p_1 = p_2 = \dots = p_N = 1/N$. En principe pour un jeu de valeurs p_i la méthode de vote, quelle qu’elle soit, désignera un et un seul vainqueur, mais le point $p_i = 1/N$ est un point de bifurcation et la moindre fluctuation peut mener au choix de n’importe lequel des candidats. Il reste que, pour M grand, cette statistique n’explore que le voisinage du point central de l’hyper-triangle.

La méthode de tirage conduisant à la statistique de B-E est plus curieuse. Son schéma est dû à Polya. Comme précédemment, l’urne contient initialement N boules. Le premier électeur choisit une boule qui indique son choix de liste et remet dans l’urne outre la boule qu’il a tirée, une autre boule portant le même numéro. Les suivants procédant de la même manière, le nombre de boules augmente d’une unité après chaque tirage.

Nous montrons dans le cas $N = M = 3$ que nous recouvrons bien la statistique de B-E. La figure 8 donne l’arborescence décrivant les tirages successifs : une branche de gauche indique le tirage d’une boule numérotée 1, une branche verticale celui d’une boule numérotée 2 et une branche de droite celui d’une boule numérotée 3. Le schéma s’explique de lui-même avec les probabilités calculées à chaque tirage qui prend en compte les tirages précédents. Les deux dernières lignes indiquent le résultat obtenu ainsi que les probabilités respectives. On récupère les 10 cas de B-E chacun avec la même probabilité $1/10$.

La statistique de F-D s’obtient encore plus simplement. Nous partons toujours de la même urne contenant les N boules numérotées de 1 à N . Il suffit, après tirage, de ne pas remettre la boule dans l’urne. La figure 9

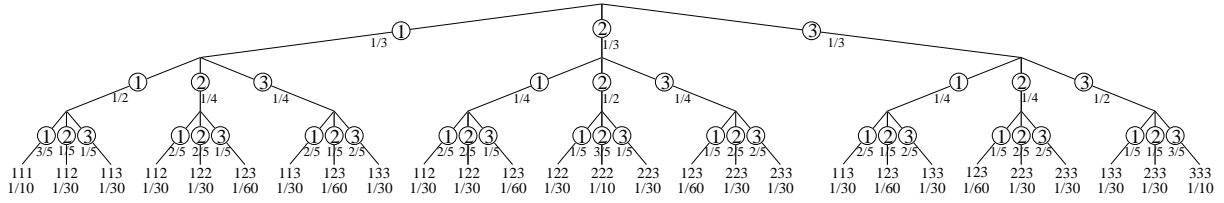


Figure 8: Arborescence de la succession des tirages dans l'urne dans le cas B-E pour $N = M = 3$. 1^{ère} ligne : nom du point obtenu (cf. figure 2), 2^{ème} ligne : probabilités respectives.

qui donne l'arborescence pour $N = 4, M = 2$, illustre l'équirépartition des points 12 13 14 23 24 34 avec pour chacun la probabilité $1/6$.

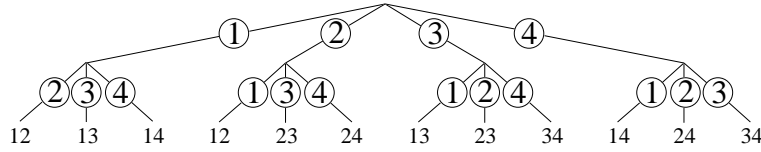


Figure 9: Arborescence de la succession des tirages dans l'urne dans le cas F-D pour $N = 4$ et $M = 2$.

Si l'on prend maintenant $N = 4$ et $M = 3$, ces 6 possibilités donnent la figure 10. Après un troisième tirage, les 4 possibilités équiprobables 123 ; 124 ; 134 ; 234 sont obtenues. Un dernier tirage amènerait 1234 de façon certaine.

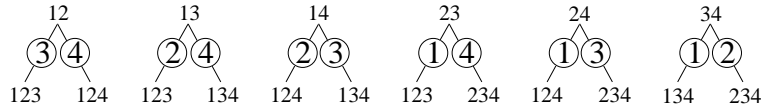


Figure 10: Dernière branche de l'arborescence dans le cas F-D pour $N = 4$ et $M = 3$.

Ainsi, l'urne de Polya contenant initialement N boules, permet d'obtenir les 3 statistiques en procédant à M tirages qui donneront respectivement :

- . la statistique de Fermi-Dirac, si on ne remet pas la boule tirée,
- . la statistique de Maxwell-Boltzmann si on remet la boule,
- . la statistique de Bose-Einstein si, en plus de la boule tirée, on en ajoute une autre identique.

Enfin, le modèle de l'urne montre que si $N \gg M$, la composition de l'urne est, dans les trois cas, peu modifiée et les trois statistiques tendent à se confondre.

5. ÉCARTS QUADRATIQUES POUR LES TROIS STATISTIQUES

Il est évidemment intéressant de voir comment varie avec N et M le volume de l'espace des phases où se trouve la majorité des points, majorité pondérée pour M-B où les schémas des figures 6 et 7 indiquent clairement la concentration des poids sur les points entourant le point central. Il est nécessaire de bien distinguer la probabilité P_i du point i de l'hyper-triangle de la fréquence p_{ij} d'apparition de la liste j au point i . L'indice j varie de 1 au nombre total de listes N , et l'indice i varie de 1 au nombre total de points de l'espace des phase donné par (3) pour M-B et B-E et par C_N^M pour F-D. Bien entendu

$$\langle p_j \rangle = \sum_i P_i p_{ij} = \frac{1}{N} \quad \forall j \quad (4)$$

par raison de symétrie. Un exemple explicite ces idées. Soit la statistique de F-D avec $N = 6$ et $M = 3$. Les points possibles, au nombre de $C_6^3 = 20$, sont listés ci-dessous

123 ; 124 ; 125 ; 126 ; 134 ; 135 ; 136 ; 145 ; 146 ; 156 ;
234 ; 235 ; 236 ; 245 ; 246 ; 256 ; 345 ; 346 ; 356 ; 456

Nous cherchons σ_1^2 l'écart quadratique de la probabilité de la liste 1 (mais par raison de symétrie tous les σ_i^2 sont égaux) :

$$\sigma_1^2 = \sum_{i=1}^{20} P_i \left(p_{i1} - \frac{1}{N} \right)^2 \quad (5)$$

Les dix premiers points listés contiennent une et une seule liste 1 sur les 3 choisies. Par conséquent, pour ces 10 points $p_{i1} = 1/3$. Quant au 10 autres, aucun des trois électeur n'a choisit la liste 1 et $p_{i1} = 0$ pour ceux là. On a donc

$$\sigma_1^2 = \frac{1}{20} \left[10 \left(\frac{1}{3} - \frac{1}{6} \right)^2 + 10 \left(0 - \frac{1}{6} \right)^2 \right] = \frac{1}{36} \quad (6)$$

Nous laissons au lecteur le soin de démontrer que pour la statistique de F-D on obtient

$$\sigma_{\text{F-D}}^2 = \frac{N - M}{MN^2} \quad (7)$$

On retrouve sur (7) la nécessité d'avoir $N \geq M$ et le fait que si $N = M$ on obtient seulement le point central $1/N 1/N \dots 1/N$ avec $\sigma_{\text{F-D}}^2 = 0$.

Le cas de M-B se traite d'une manière analogue. On obtient

$$\sigma_{\text{M-B}}^2 = \frac{N - 1}{MN^2} \quad (8)$$

Mais maintenant il n'y a pas de restriction sur les valeurs que peut prendre M et si $M \rightarrow \infty$, $\sigma_{\text{M-B}}^2 \rightarrow 0$ et la distribution devient de plus en plus piquée autour du point central. Enfin, pour $M = 1$ les relations (7) et (8) donnent le même résultat ce qui est évident sur le modèle de l'urne.

Pour le modèle B-E, l'écart quadratique est donné par

$$\sigma_{\text{B-E}}^2 = \frac{N - 1}{N + 1} \frac{N + M}{MN^2} \quad (9)$$

On note, en particulier, que les fluctuations ne décroissent pas vers zéro si $M \rightarrow \infty$ et dans cette limite $\sigma_{\text{B-E}}^2 \rightarrow (N - 1)/[N^2(N + 1)]$.

La figure 11 donne dans le cas B-E et pour $N = 3$, la trajectoire du point représentatif de l'état de l'urne au fur et à mesure du vote des électeurs. Comme il fallait s'y attendre ce sont les premiers tirages qui déterminent le point du triangle qui sera atteint après les M choix. Cependant, il n'y a pas convergence vers un point précis : simplement la trajectoire est de plus en plus "paresseuse" et le poids du passé de plus en plus lourd.

Enfin, on retrouve que les trois statistiques se confondent dans la limite $N \gg M$ soit puisque M est un entier positif, $N \rightarrow \infty$. Dans cette limite, $\sigma_{\text{B-E}}^2 = \sigma_{\text{M-B}}^2 = \sigma_{\text{F-D}}^2 \rightarrow 1/MN$.

6. SIMULATION DANS LE CAS D'UNE STATISTIQUE B-E (2^{ème} MÉTHODE)

L'utilisation de la méthode de l'urne pour étudier le cas $M \rightarrow \infty$ (opérationnellement en prenant M suffisamment grand) amènerait des calculs inutilement longs. Dans tous les cas (M fini ou infini) il est préférable

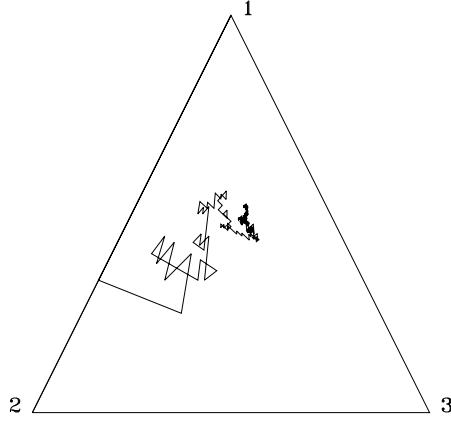


Figure 11: $N = 3$, trajectoire du point représentatif de la constitution de l'urne au fur et à mesure du vote des électeurs ($M = 1000$).

de fixer un point de l'hyper-triangle qui correspond au cas $M \rightarrow \infty$. On fixe ensuite un nouveau point, puis un troisième, etc... et c'est l'ensemble de ces points qui doit être équiréparti sur l'hyper-triangle. Enfin c'est à partir de ces points définis pour $M \rightarrow \infty$ que l'on traitera également le cas M fini. Les étapes sont les suivantes :

- On procède au tirage de $N - 1$ nombres aléatoires répartis uniformément sur l'intervalle $0 - 1$,
- On trie ces nombres aléatoires par valeur croissante. Appelons $\epsilon_1, \epsilon_2, \dots, \epsilon_{N-1}$, les nombres ainsi triés : on a $\epsilon_1 < \epsilon_2 < \dots < \epsilon_{N-1}$,
- On prend alors $p_1 = \epsilon_1, p_2 = \epsilon_2 - \epsilon_1, \dots, p_i = \epsilon_i - \epsilon_{i-1}, \dots, p_N = 1 - \epsilon_{N-1}$.

On a ainsi obtenu un point p_1, \dots, p_N tel que $\sum p_i = 1$ et qui est équiréparti sur l'hyper-triangle.

Maintenant si M est fini nous pouvons utiliser ces valeurs p_1, \dots, p_N pour déterminer le choix de chacun des M votants. Nous explicitons ce calcul pour le cas $N = M = 3$. Il permet de mieux comprendre la signification de la statistique de B-E (ou de la culture anonyme). Puisque $N = 3$, nous tirons pour commencer deux nombres aléatoires δ et ϵ dans l'intervalle $0 - 1$. Nous appelons ϵ le plus petit et δ le plus grand. Nous devons vérifier que les 10 possibilités 111, 112, 113, 122, 123, 133, 222, 223, 233, 333 sont bien équiprobables. Pour des raisons de symétrie (voir figures 2 et 3) il suffit de calculer les probabilités pour 111 112 et 123. Il est d'abord nécessaire de calculer le coefficient de normalisation à savoir

$$\int_0^1 d\delta \int_0^\delta d\epsilon = 1/2 \quad (10)$$

Le calcul pour 111 est simple. Pour ϵ et δ donné la probabilité de 111 est ϵ^3 et compte tenu du coefficient de normalisation donné par (8) la probabilité vaut

$$2 \int_0^1 d\delta \int_0^\delta \epsilon^3 d\epsilon = 1/10 \quad (11)$$

La probabilité de 112 se calcule en disant que pour ϵ et δ donnés la probabilité de 112 est $\epsilon^2(\delta - \epsilon)$ mais qu'il faut tenir compte des trois combinaisons possibles et, bien sûr, du facteur de normalisation. Ce qui donne

$$6 \int_0^1 d\delta \int_0^\delta \epsilon^2(\delta - \epsilon) d\epsilon = 1/10 \quad (12)$$

Enfin, la probabilité pour 123 est obtenue en intégrant $\epsilon(\delta - \epsilon)(1 - \delta)$ affecté du coefficient 12 (2 pour la normalisation, 6 pour les permutations).

$$12 \int_0^1 d\delta \int_0^\delta \epsilon(\delta - \epsilon)(1 - \delta)d\epsilon = 1/10 \quad (13)$$

Vérifions le bien fondé de ce calcul sur un exemple un peu plus complexe : $N = M = 4$ qui donne 35 points dans le tétraèdre et donc une probabilité de $1/35$. Le facteur de normalisation est donné par

$$\int_0^1 d\eta \int_0^\eta d\delta \int_0^\delta d\epsilon = 1/6 \quad (14)$$

où les trois nombres aléatoires fixant les 4 probabilités sont ϵ, δ, η avec $\epsilon < \delta < \eta$. Supposons que nous nous intéressions à la probabilité de 1224. Il y a 12 combinaisons menant à ce résultat. Pour une combinaison il faut calculer

$$\int_0^1 d\eta \int_0^\eta d\delta \int_0^\delta \epsilon(\delta - \epsilon)^2(1 - \eta) d\epsilon = \frac{1}{2520}.$$

Ce résultat multiplié par 6×12 redonne bien la probabilité de $1/35$.

7. RETOUR SUR LE CALCUL $M \rightarrow \infty$

Pour N listes possibles on calcule tout d'abord le facteur de normalisation : on a tiré au hasard $N - 1$ nombres que l'on ordonne

$$\epsilon_1 < \epsilon_2 < \epsilon_3 \dots < \epsilon_{N-1}$$

Le facteur de normalisation est donné par

$$\int_0^1 d\epsilon_{N-1} \int_0^{\epsilon_{N-1}} d\epsilon_{N-2} \dots \int_0^{\epsilon_2} d\epsilon_1 = \frac{1}{(N-1)!} \quad (15)$$

et la valeur moyenne d'une quantité f est donnée par :

$$(N-1)! \int_0^1 d\epsilon_{N-1} \int_0^{\epsilon_{N-1}} d\epsilon_{N-2} \dots \int_0^{\epsilon_2} f d\epsilon_1 \quad (16)$$

On calcule ainsi la valeur moyenne de $p_1 = \epsilon_1 = f$. On obtient $\langle p_1 \rangle = 1/N$ comme il fallait s'y attendre. En fait $\langle p_1 \rangle = \langle p_2 \rangle = \dots = \langle p_N \rangle = 1/N$. L'écart quadratique moyen de p_1 est donné par

$$(N-1)! \int_0^1 d\epsilon_{N-1} \int_0^{\epsilon_{N-1}} d\epsilon_{N-2} \dots \int_0^{\epsilon_2} (\epsilon_1 - \frac{1}{N})^2 d\epsilon_1 = \frac{N-1}{N^2(N+1)} \quad (17)$$

En fait

$$\langle (p_i - \frac{1}{N})^2 \rangle = \frac{N-1}{N^2(N+1)} \quad (18)$$

Ce dernier résultat est intéressant puisque nous retrouvons par ce calcul qui traite directement le cas $M \rightarrow \infty$ celui obtenu en travaillant sur les positions des points discrétisés de l'hyper-triangle (voir partie 5). Il confirme l'intérêt de ce deuxième modèle pour les larges électorats.

Les formules (17) et (18) recourent un résultat connu en théorie des probabilités. Si un nombre infini de points sont répartis au hasard sur une droite avec une densité n (c'est-à-dire si N points sont répartis uniformément sur un segment L et si N et L tendent vers l'infini (avec $n = N/L$ constant), alors les intervalles x entre deux points tendent vers une répartition de Poisson de paramètre n . De plus la caractéristique d'une loi de Poisson est que l'écart quadratique moyen est le carré de la moyenne. Ici pour $N \rightarrow \infty$ on voit qu'en effet l'écart quadratique moyen est $1/N^2$ cependant que la moyenne est $1/N$.

On peut vérifier en effet que si N est suffisamment grand (pratiquement à partir de 10), les p_i suivent la loi de Poisson $N \exp -Np_i$. Cela n'est évidemment pas vrai pour des N petits. Par exemple pour $N = 2$, p_1 et p_2 sont équirépartis dans l'intervalle $0 - 1$. Il est d'ailleurs possible de calculer la probabilité d'obtenir la

probabilité p_1 . Par exemple pour $N = 4$, nous tirons 3 nombres $\gamma, \delta, \varepsilon$, avec $\gamma > \delta > \varepsilon$ et $p_1 = \varepsilon$. La probabilité que p_1 soit supérieur à une valeur a de l'intervalle $0 - 1$ est

$$6 \int_a^1 d\gamma \int_a^\gamma d\delta \int_a^\delta d\varepsilon = (1 - a)^3 \quad (19)$$

Ce résultat se généralise aisément et donne pour N listes

$$Prob(p_1 > a) = (1 - a)^N \quad (20)$$

La relation (20) doit être comparé à la limite Poissonnienne $Prob(p_1 > a) = \exp(-Na)$.

Curieusement on peut, quelque soit N , utiliser la formule de Poisson pour tirer les N probabilités p_i et récupérer une distribution identique à celle obtenue par le tri de $N-1$ nombre pris dans l'intervalle $0-1$ (cf. partie 6). Ce résultat étrange à première vue vient de la nécessité de normaliser les nombres tirés de la Poissonnienne tandis que la méthode du tri assure automatiquement cette normalisation. Si donc on a tiré p_1, \dots, p_n à partir d'une distribution de Poisson, il faut en fait considérer $\bar{p}_1 = p_1 / \sum_i p_i, \dots, \bar{p}_N = p_N / \sum_i p_i$. La figure 12 explicite la construction dans le cas $N = 2$ (pour lequel nous devons transformer la Poissonnienne $2 \exp(-2p_1)$ en distribution uniforme sur l'intervalle $0-1$). Si p_1 et p_2 sont les valeurs issues du tirage Poissonnien, les valeurs normalisées sont obtenues à l'intersection de OM et du segment joignant les points $(0, 1)$ et $(1, 0)$. Appelons A le point d'abscisse a situé sur ce segment, la probabilité d'avoir $\bar{p}_1 < a$ est donnée par la probabilité d'avoir M dans l'angle formé par OA et l'axe p_2 . Ce calcul est effectué en considérant l'élément de surface hachuré compris entre l'axe p_2 , la droite OA et les deux droites $p_1 + p_2 = \xi$ et $p_1 + p_2 = \xi + d\xi$. La longueur du segment BC est $a\sqrt{2}\xi$ et la hauteur du trapèze hachuré est $d\xi/\sqrt{2}$. La probabilité cherchée s'écrit :

$$\int_0^\infty \frac{d\xi}{\sqrt{2}} a\sqrt{2}\xi 4 \exp(-2\xi) = a \quad (21)$$

donnant une probabilité $1 - a$ que $p_1 > a$, en accord avec (20).

Le cas $N = 3$ est encore justifiable d'une interprétation géométrique dans un espace à 3 dimensions (voir fig 13). Le plan CDE est défini par $p_3 = a$. Pour que p_3 soit inférieur à a , le point (p_1, p_2, p_3) doit se trouver à l'intérieur de la pyramide définie par les quatre droites OA, OB, OC et OD . Le trapèze $ABCD$ a pour surface $a\sqrt{3}(1 - a/2)$. On définit l'élément de volume entre les plans $p_1 + p_2 + p_3 = \xi$ et $p_1 + p_2 + p_3 = \xi + d\xi$, sa hauteur est $d\xi/\sqrt{3}$. On doit alors calculer :

$$\int_0^\infty \frac{d\xi}{\sqrt{3}} a\sqrt{3}(1 - \frac{a}{2})\xi^2 3^3 \exp(-3\xi) = 2a - a^2 \quad (22)$$

donnant une probabilité $1 - 2a + a^2 = (1 - a)^2$ que $p_1 > a$, en accord avec (20).

Le tirage dans la distribution de Poisson des probabilités des N listes, suivie de la renormalisation appropriée est une alternative conceptuellement plus simple que le tirage et le tri des $N-1$ nombres aléatoires et si les deux algorithmes sont de rapidité comparables, la méthodes de la Poissonnienne est beaucoup moins gourmande en place mémoire que celle du tri.

8. STATISTIQUE DE M-B (CULTURE IDENTIFIABLE), CAS DE $M \rightarrow \infty$

La statistique M-B dans la limite où $M \rightarrow \infty$ concentre la quasi totalité des poids dans une zone centrée autour du point $(\frac{1}{N}, \dots, \frac{1}{N})$ et dont la taille décroît comme $M^{-1/2}$. Contrairement au cas B-E où nous explorons l'hypertriangle, nous restons dans le voisinage d'un point. En revanche ce point est très particulier et son voisinage aussi petit qu'il soit (c'est-à-dire aussi grand que soit M) présente une variété de situations analogues à celles rencontrées dans l'exploration de l'hypertriangle de B-E. La raison est la suivante. Supposons $M \rightarrow \infty$ on aura pour p_i la valeur $1/N + \epsilon\alpha_i$ où ϵ est un nombre qui décrit la taille des fluctuations (il varie donc en $M^{-1/2}$) et α_i précise cette fluctuation. Considérons pour fixer les idées le duel AB dans le cas de trois candidats. Son issue dépend des valeurs respectives de $p_1 + p_2 + p_5$ et de $p_3 + p_4 + p_6$ c'est-à-dire de

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \epsilon(\alpha_1 + \alpha_2 + \alpha_5) \text{ et de } \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \epsilon(\alpha_3 + \alpha_4 + \alpha_6) \quad (23)$$

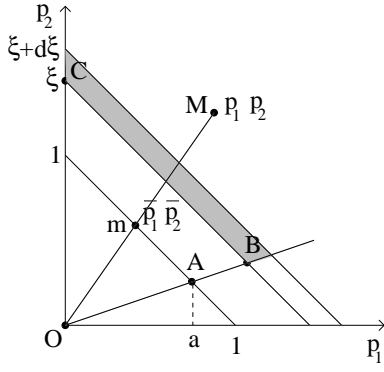


Figure 12:

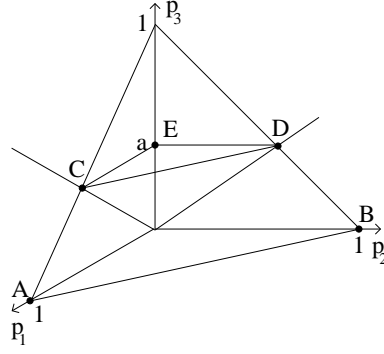


Figure 13:

liste n°	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	C	D	D	D	D	D	D
	B	B	C	C	D	D	A	A	C	C	D	D	A	A	B	B	D	D	A	A	B	B	C	C
	C	D	B	D	B	C	C	D	A	D	A	C	B	D	A	D	A	B	B	C	A	C	A	B
	D	C	D	B	C	B	D	C	D	A	C	A	D	B	D	A	B	A	C	B	C	A	B	A

Table 2: Énumération des $N = 24$ listes dans le cas de $n = 4$ candidats notés A, B, C et D.

On voit que la valeur de ϵ ne joue aucun rôle et que seules comptent les valeurs relatives de $\alpha_1 + \alpha_2 + \alpha_5$ et de $\alpha_3 + \alpha_4 + \alpha_6$. On vérifie qu'il en est de même non seulement pour tous les duels mais également dans toutes les comparaisons de points liés à l'emploi de règles positionnelles. Pour réaliser une simulation dans la limite $M \rightarrow \infty$ il suffira de se rappeler que l'on peut utiliser le théorème centrale limite et que, par conséquent, les α_i seront tirés au hasard et indépendemment les uns des autres dans la distribution Gaussienne.

$$p(\alpha) = \frac{1}{\sqrt{2\pi}} \exp -\frac{\alpha^2}{2} \quad (24)$$

Si on se rappelle que l'on avait normalisé avec $\sum p_i = 1$ on doit procéder à la correction

$$p_i = \frac{\bar{p}_i}{\sum \bar{p}_i} \quad (25)$$

où

$$\bar{p}_i = \frac{1}{N} + \epsilon \bar{\alpha}_i. \quad (26)$$

En tenant compte du fait que $\epsilon \rightarrow 0$ il vient

$$p_i = \frac{1}{N} + \epsilon \left(\bar{\alpha}_i - \frac{1}{N} \sum \bar{\alpha}_i \right). \quad (27)$$

On voit donc que, une fois tiré les $N \bar{\alpha}_i$ dans une répartition Gaussienne, il suffit de les recentrer de telle sorte que $\sum \alpha_i = 0$ avec $\alpha_i = \bar{\alpha}_i - (1/N) \sum \bar{\alpha}_i$, on remarquera l'analogie avec le cas précédent : les N valeurs décrivant la situation sont maintenant tirées dans une distribution de Gauss au lieu d'une distribution de Poisson.

Pour terminer ces considérations sur la technique de simulation mentionnons la solution d'un problème de taille mémoire d'ordinateur. Si n est grand (et il n'a pas besoin d'être bien grand), il faut considérer $N = n!$ listes de n noms. La saturation vient rapidement. L'idée est alors de donner un numéro à ces listes permettant de retrouver rapidement sa structure. A titre d'exemple, prenons $n = 4$ ($N = 24$). La table (2) donne les listes classées par ordre alphabétique de 0 à 23.

Notons tout d'abord la structure auto-similaire de cette liste. Si on retire la première ligne, chaque bloc de 6 listes (listes de 0 à 5, de 6 à 11 etc...) est l'ensemble des listes obtenues pour 3 candidats pris dans la liste

B,C,D ou A,C,D etc.. (le candidat de la première ligne A, B, etc... a été retiré de la liste A,B,C,D). De même en retirant la deuxième ligne et en considérant les listes par bloc de 2, on se retrouve avec les listes à 2 candidats.

L'ordre alphabétique, choisi ici, permet d'élaborer l'algorithme suivant : supposons que nous désirions connaître la structure de la liste 16. On divise ce nombre par $(n - 1)!$ (ici $3! = 6$), on a

$$16 = 2 \times 6 + 4$$

Le quotient - forcément inférieur à n - donne le nom du candidat placé en première position. Pour cela on choisit le candidat occupant le rang quotient (ici 2) dans la liste ABCD renumérotée 0 1 2 3, c'est donc C. On retire C de la liste qui comporte maintenant A,B,D de rang 0,1,2 respectivement. On réitère en prenant le reste que l'on divisera par $(n - 2)!$ Ici il vient

$$4 = 2 \times 2 + 0$$

Le nouveau quotient 2 désigne maintenant D dans la liste ABD. Le nouveau reste 0 implique pour le troisième quotient 0 et donc le choix A dans AB. Enfin on obtient B qui indique CDAB pour la liste numéro 16 en accord avec la table (2).

L'algorithme inverse, trouver le numéro d'une liste donnée, ne pose, lui-aussi, aucun problème.

L'un des apports de ce papier est de donner l'éclairage de la physique sur l'aspect statistique de la théorie des votes. Ainsi, en poursuivant l'ébauche de dictionnaire de la partie 3, il est possible d'apporter quelques commentaires sur la réalité des différents modèles qui permettent de se situer dans les distributions B-E, M-B ou F-D, à savoir le tirage dans l'urne ou le tri de $N - 1$ nombres aléatoires comme indiqué partie 6. Cette discussion est assez peu abordée, toutefois, parmi d'autres, deux articles [2] [10] avancent des explications basées sur le modèle de l'urne (cf section 4) généralisée en ajoutant à chaque tirage non pas une boule, mais α boules. B-E apparaît alors comme le cas particulier $\alpha = 1$ et F-D comme le cas $\alpha = -1$. Grâce à ce degré de liberté supplémentaire, les auteurs interprètent la "culture anonyme" (généralisée avec $\alpha > 1$) comme correspondant à un degré d'homogénéité social. Tiré de [10], nous citons : "*Clearly α is a contagion parameter, a measure of the mutual influence between voters ... then increasing α will increase the propensity of voters to have similar preference and will reflect increase social homogeneity*". Conséquemment, la "culture identifiable" (M-B) pour laquelle α égale 0 refléterait alors l'indépendance des électeurs.

Certainement la "contagion" liée au modèle de l'urne est d'une nature très étrange. Ainsi, c'est le vote des premiers électeurs qui va déterminer le pourcentage des listes et pour α tendant vers l'infini, on peut même vérifier que l'issue du scrutin est imposée par le seul premier électeur. La connection avec l'idée d'homogénéité du corps électoral est alors assez délicate à interpréter.

Afin de mettre les principes de la physique à contribution, il est préférable de s'en tenir à la description des cultures en termes de statistiques, ces dernières ayant l'avantage d'être reliées à des modèles précis. En physique, ces statistiques permettent notamment de bien clarifier la façon dont on doit considérer les votes pour les deux cultures : elles sont établies en supposant des particules sans interaction (c'est-à-dire, en langage électoral, des électeurs votants de façon indépendantes). L'influence entre particules (électeurs) traduite par la présence de forces attractrices ou répulsives crée des corrélations qui sont indépendantes des effets associés aux statistiques quantiques. Ainsi la condensation de Bose-Einstein, pour laquelle un grand nombre de particules tombent sur le même niveau d'énergie (beaucoup d'électeurs choisissent la même liste) n'a pas pour origine les interactions entre particules (influence entre électeurs) mais le cadre mathématique de la mécanique quantique.

Malgré un traitement similaire, les deux cultures "anonyme" et "identifiables" associées aux statistiques de B-E et M-B sont de nature très différentes. La statistique de M-B est liée au cas classique pour lequel les phénomènes quantiques ne se manifeste pas et M électeurs indépendants choisissent chacun au hasard l'une des N listes de manière équiprobable. Pour un large électorat, le point de l'espace des phases se place très près du point central qui est un point hypercritique : un faible changement dans le choix des électeurs peut faire basculer l'issue du scrutin.

Maintenant, si une partie des électeurs vote délibérément de façon identique, les listes ne sont plus équiprobables. Le point quitte alors la région du point critique et l'issue du scrutin n'a plus l'extrême sensibilité aux fluctuations

A	A	B	B	C	C
B	C	A	C	A	B
C	B	C	A	B	A
p_1	p_2	p_3	p_4	p_5	p_6

Table 3: Les $N = 6$ listes établies pour $n = 3$ candidats notés A, B, C avec leurs probabilités p_i respectives.

comme précédemment.

La distribution de B-E, quant-à elle, est de nature quantique et fait donc appel à des arguments probabilistes. Il n'est plus possible de localiser le système en un seul point de l'espace des phases électoral. Il faut maintenant réaliser une moyenne sur plusieurs scrutins afin d'obtenir une information pertinente. C'est bien ce qui a été fait lors des simulations numériques et cela appelle les remarques suivantes :

- Pour chaque scrutin (qui donne un point de l'espace des phases) les électeurs votent indépendamment les uns des autres.
- Pour un scrutin, les probabilités des listes (prise dans la distribution de Poisson) peuvent-être notablement différentes.
- Cependant, la moyenne des probabilités du choix d'une liste i sur l'ensemble des scrutins donne bien $\langle p_i \rangle = 1/N$ comme calculé partie 5.

On voit que les remarques précédentes sont très liées à l'utilisation du modèle introduit partie 6 (tirage et tri de nombres aléatoires) et, bien sûr, à la "culture physique" des auteurs.

Encore une fois, pour tenir compte de l'éventuelle présence d'une homogénéité sociale, il faudrait biaiser le tirage dans la Poissonnienne afin de faire apparaître un effet systématique pour la liste privilégiée.

9. INFLUENCE DU NOMBRE DE CANDIDATS SUR L'EFFET CONDORCET

La méthode préconisée par Condorcet est une généralisation des duels binaires à partir des listes fournies par les électeurs. Montrons sur 3 candidats comment procéder et quels sont les problèmes rencontrés. Soit A, B, C les candidats. Les 6 listes sont indiquées tableau 3 (avec sur la ligne du haut le candidat préféré) ainsi que les pourcentages obtenus. Trois duels sont à envisager : le résultat du duel AB est décidé en comparant $p_1 + p_2 + p_5$ avec $p_3 + p_4 + p_6$. De même AC implique la comparaison de $p_1 + p_2 + p_3$ et de $p_4 + p_5 + p_6$. cependant que le duel BC a son issue décidée en comparant $p_1 + p_3 + p_4$ et $p_2 + p_5 + p_6$. Se peut-il alors qu'un candidat gagne tous ses duel ? Il serait alors normal de le déclarer élu (encore que d'autres méthodes d'agrégation *a priori* aussi "légitime" que celle de Condorcet peuvent amener un autre prétendant à la victoire). Mais déjà, en se restreignant à la méthode de Condorcet, des difficultés apparaissent avec, par exemple, un résultat cyclique où A bat B , B bat C et C bat A . On vérifiera que c'est le cas, avec pour chaque duel un score de 52% au vainqueur contre 48%, pour le jeu de valeurs : $p_1 = .42, p_2 = .09, p_3 = .07, p_4 = .13, p_5 = .11, p_6 = .28$. Ce résultat, un peu surprenant de prime abord, n'a rien de vraiment paradoxal et nous préférons la dénomination "d'effet Condorcet" à celle de "paradoxe de Condorcet".

Dans le cas $n = 3$ ($N = 6$), la fréquence de "l'effet Condorcet" pour un nombre M impair et fini de votants a pu être calculé analytiquement pour une statistique du type B-E (culture anonyme) [7]. La probabilité qu'un vainqueur existe est donnée par

$$P(3, M) = \frac{15}{16} \frac{(M + 3)^2}{(M + 2)(M + 4)}, \quad (28)$$

avec $M = 1, 3, 5, \dots$. On vérifie que $M = 1$ donne $P(3, 1) = 1$ (un seul votant, ou dictateur simplifie toujours le processus électoral !).

Nous avons procédé à l'étude des cas $n = 3, 4, 5, 6$ pour différentes valeurs de M et à celles des cas $n = 7$ et $n = 8$ pour M infini. Rappelons que pour $n = 3$ (28) donne l'expression analytique dans le cas d'une statistique B-E et que pour $n = 3$ et 4 et $M = 3, 5, 7$ les valeurs exactes ont été obtenues par énumérations et examen de tous les points de l'espace des phases électoral (un peu plus de 2×10^6 points pour le cas $M = 7, n = 4$).

Les résultats sont présentés sur les figures 14 (culture anonyme) et 15 (culture identifiable). Sur la figure 14 et pour $n = 3$, on a tracé en pointillé la courbe théorique donnée par (28). On notera l'excellent accord y compris pour la valeur théorique 0.0625 (soulignons que nous avons porté la probabilité \bar{P} de ne pas avoir de vainqueur Condorcet). La comparaison des résultats pour $n = 3$ et $n = 4$ dans le cas des cultures identifiables confirme la relation de May $\bar{P}(4, M) = 2\bar{P}(3, M)$. En revanche, il est hors de doute que cette relation n'est pas valable dans le cas d'une culture anonyme contrairement à ce qui est affirmé en [7]. Notons que ce désaccord avait déjà été souligné en [1].

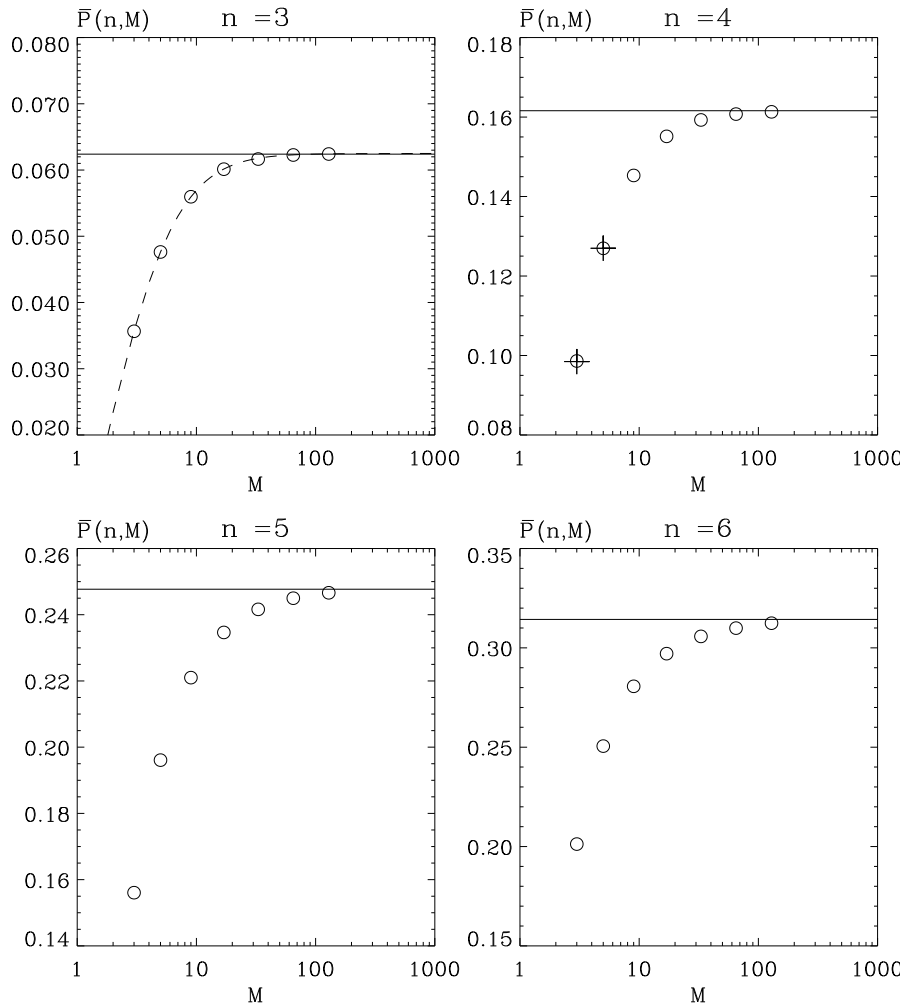


Figure 14: Probabilité $\bar{P}(n, M)$ de conflits Condorcet dans le cas d'une culture anonyme (statistique B-E), pour un nombre M fini de votants et pour $n = 3, 4, 5, 6$. Les croix correspondent aux valeurs théoriques, la droite est la valeur obtenue pour $M \rightarrow \infty$.

Les tables 4 et 5 donnent $P(n, \infty)$ pour n variant de 3 à 8 et pour les deux "cultures". Un point intéressant,

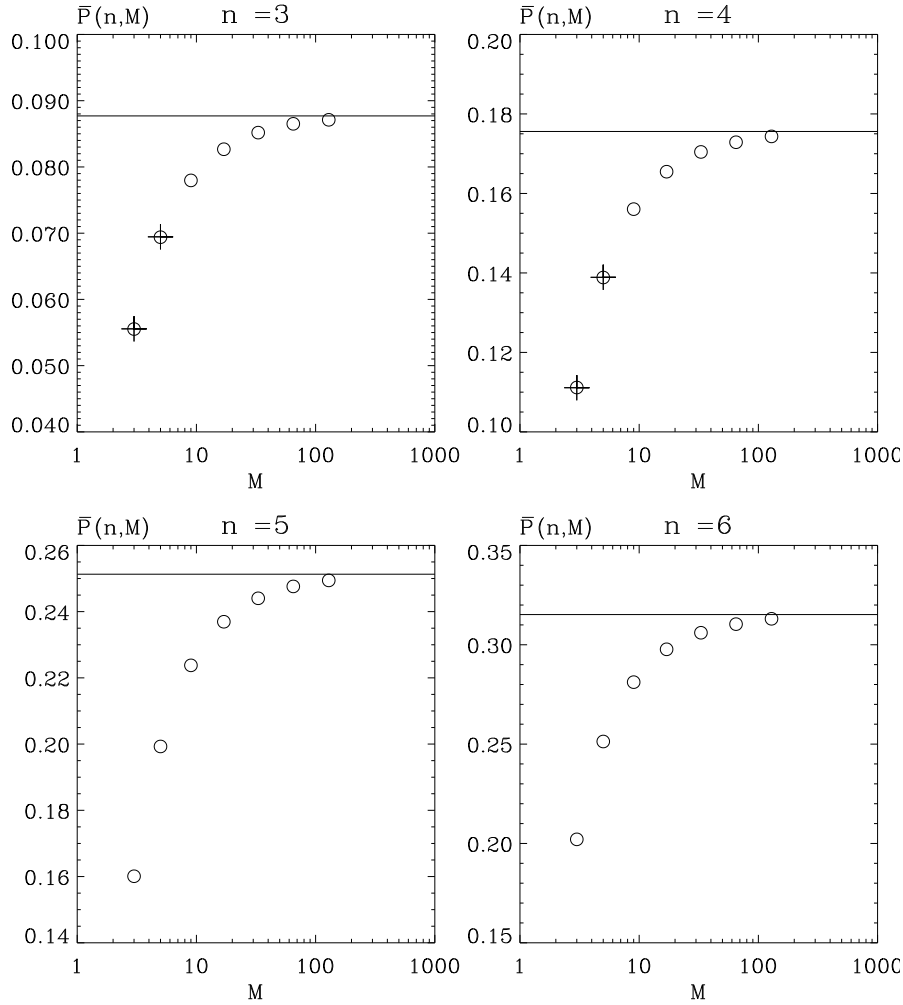


Figure 15: Probabilité $\bar{P}(n, M)$ de conflits Condorcet dans le cas d'une culture identifiable (statistique M-B), pour un nombre M fini de votants et pour $n = 3, 4, 5, 6$. Les croix correspondent aux valeurs théoriques, la droite est la valeur obtenue pour $M \rightarrow \infty$.

prédit en [1]. est la pratique identité des résultats relatifs aux deux cultures dès que $n \geq 6$. Ceci mérite un commentaire. On a vu que la statistique B-E entraînait pour la distribution des p_i (la probabilité d'associer la probabilité p_i à la liste i) une loi de Poisson de moyenne $\frac{1}{N}$ et de variance $\frac{1}{N^2}$ dans la limite $N \rightarrow \infty$. L'utilisation de la statistique de M-B introduit pour p_i une répartition Gaussienne de valeur moyenne $1/N$ et dont l'écart type décroît comme $M^{-1/2}$. Donc, a priori les deux statistiques sont différentes. L'identité des résultats finaux sur $P(n, \infty)$ avec $n! \rightarrow \infty$ vient, une fois de plus, du théorème central limite. L'issue d'un duel dépend de la comparaison de deux sommes comprenant chacune si $N \rightarrow \infty$ un grand nombre de p_i . Dans les deux cas on obtiendra pour ces sommes des distributions Gaussiennes.

On notera enfin que la précision obtenue est d'environ 10^{-3} et parfois meilleure. Il a fallu pour cela procéder à 10^7 expériences. Les calculs a priori les plus longs concernent le cas B-E où il faut à chaque expérience trier $N - 1$ nombres (si l'on renonce à tirer dans la Poissonnienne). Il est alors nécessaire d'utiliser les techniques de tri rapide pour retrouver des temps de calculs comparables dans tous les cas.

n	3	4	5	6	7	8
$P(n, \infty)$.9376	.8384	.7523	.6857	.6309	.583

Table 4: Probabilité $P(n, M = \infty)$ de vainqueur Condorcet dans le cas d’une culture anonyme (statistique B-E), pour un nombre M infini de votants.

n	3	4	5	6	7	8
$P(n, \infty)$.9123	.8244	.7487	.6848	.6306	.586

Table 5: Probabilité $P(n, M = \infty)$ de vainqueur Condorcet dans le cas d’une culture identifiable (statistique M-B), pour un nombre M infini de votants.

La méthode de tri rapide de N réels procède d’une variante du tri des N premiers entiers par adressage [9]. Dans cette dernière méthode, le pointeur de la mémoire place le nombre à trier dans la case mémoire dont l’adresse est donnée par le nombre lui-même. Pour étendre cette méthode au tri de N réels x_i répartis dans l’intervalle $[0, 1[$ par exemple, on dispose d’un vecteur de k cases, avec $k > N$. On appellera ce vecteur table rapide. Ce vecteur représente le segment $[0, 1[$ divisé en k intervalles égaux. Trier le nombre x_i , revient à l’attribuer à la case numéro $\text{Ent}(kx_i) + 1$ de la table rapide (où $\text{Ent}(x)$ désigne la partie entière de x). Pour limiter les cas pour lesquels deux nombres proches sont attribués à la même case, on pourrait prendre des tables telles que $k \gg N$. Cependant, si un conflit arrive, on le règle en procédant à un décalage en chaîne du plus grand des nombres vers une case vide de la table rapide. Pour obtenir un bon compromis entre prendre une table grande, mais quasi-vide et une table courte, mais sujette à de nombreux conflits, on adopte $k = 4N$.

10. CONCLUSION

L’espace des phases électoral a pour dimension $N = n!$, où n est le nombre de candidats. Dès que $n \geq 4$, tout traitement (et même toute collecte de l’information) fait nécessairement appel au vote électronique.

Trois modèles ont été présentés. Le premier inspiré du comptage qui en physique statistique mène à la distribution de Maxwell-Boltzmann (qualifié de “culture identifiable”), explore, pour un très large électorat, le voisinage du point hyper-critique $p_1 = \dots = p_i = \dots = p_N = 1/N$. En ce sens, il n’est pas vraiment représentatif d’une distribution. Le deuxième associé à la statistique de Bose-Einstein (qualifié de “culture anonyme”) correspond, même pour un électorat infini, à l’exploration de tout le volume accessible de l’espace des phases. Le troisième associé à la statistique de Fermi-Dirac n’est pas utilisé en théorie des votes, mais pourrait avoir un intérêt dans des problèmes de choix unique.

Pour simuler ces deux cultures (ou en physique ces trois statistiques) deux modèles ont été décrits. Le premier est basé sur un tirage de boules dans une urne dont la composition varie au cours du tirage pour les statistiques de Fermi-Dirac et Bose-Einstein. La seconde méthode s’applique au cas $M \rightarrow \infty$ et est basée sur le tri de $N - 1$ nombres tirés au hasard dans l’intervalle $[0, 1]$ ou, alternativement, sur des tirages dans la distribution de Poisson ou de Gauss.

Les résultats obtenus pour n variant de 3 à 8 confirment et précisent ceux obtenus et discuté auparavant à savoir la validité de la formule de May $\bar{P}(4, M) = 2\bar{P}(3, M)$ mais uniquement dans le cas des cultures identifiables et l’égalité de $\bar{P}(n, \infty)$ pour les deux cultures dès que $n \geq 6$. Enfin, la probabilité d’avoir un “vainqueur Condorcet” décroît régulièrement quand n augmente mais reste encore égale à 58% quand $n = 8$ (pour les deux cultures et pour un électorat infini).

Il existe d’autres modèles. Dans l’un d’entre eux, le nombre M_1 d’électeurs ayant choisi la liste 1, est tiré au hasard parmi les entiers $0, 1, \dots, \mathcal{N}$. De même, on détermine les nombres $M_2, \dots, M_{\mathcal{N}}$. On remarquera, que, dans cette approche, M n’est pas fixé (comme pour les ensembles grands canoniques). Dans la limite ou $\mathcal{N} \rightarrow \infty$, on obtient les pourcentages de la liste i en tirant au hasard ϵ_i et en normalisant $p_i = \epsilon_i / \sum_i \epsilon_i$. L’obtention des p_i est, dans ce cas, particulièrement simple.

REMERCIEMENTS

L'un des auteurs (M.R.F.) remercie Dominique Lepelley et Vincent Merlin pour l'avoir introduit à ce sujet ainsi que pour une intéressante discussion sur l'interprétation des deux cultures.

BIBLIOGRAPHIE

- [1] Berg S. and Bjurulf B., “A note on the paradox voting ; anonymous preference profiles and Mary’s formula”, *Public Choice*, 40 (1983), 307–316.
- [2] Berg S. et Lepelley D., “Note sur le calcul de la probabilité des paradoxes du vote”, *Mathématique, Informatique et Sciences humaine*, 120 (1992), 33–48.
- [3] Black D., *The theory of committee and elections*, Cambridge University Press, Cambridge, 1958.
- [4] Borda J.C., *Mémoires sur les élections au scrutin*, Histoires de l’Académie Royale des Sciences, Paris, 1781.
- [5] Condorcet M.J.A., *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des votes*, Paris, 1785.
- [6] Feller W., *An introduction to probability theory and its applications*, 2nd edition, New-York, Wiley, 1957.
- [7] Gehrlein W.V. and Fishburn P.C., “Condorcet’s paradox and anonymous preference profiles”, *Public Choice*, 26 (1976), 1–18.
- [8] Huang K., *Statistical mechanics*, New-York, Wiley, 1958.
- [9] Knuth D.E., *The art of computer programming: Vol. 3, sorting and searching* Reading, MA, Addison-Wesley, 1973.
- [10] Lepelley D., Pierron P. and Valognes F., “Scoring Rules, Condorcet Efficiency and Social Homogeneity”, Centre de Recherches en Économie Mathématique et Économétrie, Cahier de Recherche 99–6 (1999), 1–22.
- [11] Saari D.G., *Basic geometry of voting*, Heidelberg, Springer, 1995.