



## Procédure de vote parallèle dans les référendums multiples : une approche expérimentale

Helene Fargier, Jérôme Lang, Jérôme Mengin, Nicolas Schmidt

### ► To cite this version:

Helene Fargier, Jérôme Lang, Jérôme Mengin, Nicolas Schmidt. Procédure de vote parallèle dans les référendums multiples : une approche expérimentale. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3, 2012. <hal-00656520>

**HAL Id: hal-00656520**

**<https://hal.archives-ouvertes.fr/hal-00656520>**

Submitted on 17 Jan 2012

**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.

# Procédures de vote parallèle dans les référendums multiples : une approche expérimentale

H. Fargier<sup>1</sup>

J. Lang<sup>2</sup>

J. Mengin<sup>1</sup>

N. Schmidt<sup>1</sup>

<sup>1</sup> IRIT

<sup>2</sup> LAMSADE

118 route de Narbonne, 31062 Toulouse Cedex  
{Fargier,Mengin,Schmidt}@irit.fr ; Lang@lamsade.dauphine.fr

## Résumé

Les référendums multiples consistent à prendre une décision commune sur chacune d'un ensemble de questions binaires, à partir des préférences d'un ensemble de votants. Demander aux votants leur avis sur toutes les combinaisons de valeurs est pratiquement infaisable, en raison du nombre exponentiellement grand de ces combinaisons ; d'un autre côté, effectuer des votes en parallèle sur chacune des questions peut mener à des résultats fortement paradoxaux. Dans cet article, nous essayons de mesurer à quel point il est sous-optimal de procéder à un tel vote parallèle, en fonction de la règle de vote que l'on veut implémenter, et de la nature des préférences des votants (arbitraires, faiblement séparables ou fortement séparables). Nous utilisons pour cela une expérimentation sur des données générées aléatoirement.

## Mots Clef

Choix social computationnel ; préférences ; vote ; domaines combinatoires ; génération aléatoire.

## Abstract

Multiple referenda consists in making a common decision about each of a set of binary questions, from the preferences of a set of voters. Asking voters for their preferences on all combinations of values is not feasible in practice, because of the exponentially large number of such combinations ; on the other hand, voting in parallel on each of the questions can lead to strongly paradoxical outcomes. In this paper we try to measure to which extent it is suboptimal to proceed to parallel voting, in function of the voting rule to be implemented, and the nature of the voters' preferences (arbitrary, separable or additively separable). For this we use randomly generated profiles.

## Keywords

Computational social choice ; preferences ; voting ; combinatorial domains ; random generation.

## 1 Introduction

Dans de nombreux contextes pratiques de décision collective, les votants doivent se mettre d'accord sur une valeur à donner à chacune d'un ensemble de questions sur lesquelles ils peuvent avoir des dépendances préférentielles. Par exemple, les membres d'une association doivent élire un président, un vice-président et un secrétaire. Les votants ont typiquement des dépendances préférentielles sur ces questions, c'est-à-dire, par exemple, que les préférences sur le vice-président et le secrétaire peuvent dépendre du choix du président (par exemple, on ne souhaite pas que le président et le vice-président soient des amis intimes).

Une approche simple consiste à décomposer le vote en un ensemble de votes élémentaires, effectués simultanément - c'est-à-dire en parallèle -, question par question. A partir du moment où les votants ont des dépendances préférentielles, cette décomposition en votes élémentaires risque en théorie de conduire à des résultats contre-intuitifs [5, 10, 2, 12]. Considérons un premier exemple avec cinq votants et deux questions booléennes  $A$  et  $B$ , dont les domaines de valeurs sont respectivement  $\{a, \bar{a}\}$  et  $\{b, \bar{b}\}$ . Les préférences des votants sont comme suit :

2 votants :  $a\bar{b} \succ \bar{a}b \succ \bar{a}\bar{b} \succ ab$

2 votants :  $\bar{a}b \succ a\bar{b} \succ \bar{a}\bar{b} \succ ab$

1 votant :  $ab \succ a\bar{b} \succ \bar{a}b \succ \bar{a}\bar{b}$

On constate que le dernier votant préfère  $a$  à  $\bar{a}$  quelle que soit la valeur (fixée) de  $B$  ; et *vice versa* : il a des préférences *séparables*. En revanche, les préférences des 4 autres votants ne sont pas séparables : par exemple, les deux premiers votants préfèrent  $A = a$  à  $A = \bar{a}$  si  $B = \bar{b}$  et préfèrent  $A = \bar{a}$  à  $A = a$  si  $B = b$ . Si on demande aux votants de voter séparément, en parallèle, sur chacune des variables, et s'ils adoptent un comportement optimiste (les deux premiers votants choisissent  $A = a$  plutôt que  $A = \bar{a}$ ), ce qui est souvent le cas en pratique, alors on aboutira à une majorité pour  $A = a$  et une majorité pour  $B = b$ , d'où la décision commune  $(a, b)$ , qui est la pire pour quatre des cinq votants !

Comme le soulignent Lacy et Niou [10], le vote ques-

tion par question en parallèle (par souci de concision on dira maintenant “vote en parallèle”) est bien moins problématique lorsque les votants ont des préférences séparables<sup>1</sup>. Cela dit, la séparabilité, même dans sa version la plus faible, est une condition extrêmement exigeante, et par ailleurs, même dans sa version la plus forte, ne permet pas d’éviter tous les paradoxes, comme le montre l’exemple suivant, avec trois votants et trois questions booléennes  $A$ ,  $B$  et  $C$ . Les préférences des votants sont comme suit :

$v_1 : \bar{a}\bar{b}\bar{c} \succ \bar{a}\bar{b}c \succ \bar{a}b\bar{c} \succ \bar{a}bc \succ abc \succ a\bar{b}\bar{c} \succ a\bar{b}c \succ ab\bar{c}$   
 $v_2 : \bar{a}\bar{b}c \succ \bar{a}b\bar{c} \succ \bar{a}bc \succ \bar{a}bc \succ abc \succ a\bar{b}\bar{c} \succ ab\bar{c} \succ ab\bar{c}$   
 $v_3 : \bar{a}bc \succ \bar{a}b\bar{c} \succ \bar{a}bc \succ \bar{a}bc \succ abc \succ ab\bar{c} \succ a\bar{b}\bar{c} \succ ab\bar{c}$

Ces trois votants ont des préférences séparables. Le vote parallèle conduit à la décision collective ( $A = a, B = b, C = c$ ), qui est dominée au sens de Pareto par la décision ( $A = \bar{a}, B = \bar{b}, C = \bar{c}$ ) : en effet, tous les votants préfèrent  $\bar{a}\bar{b}\bar{c}$  à  $abc$ . Ceci met en évidence la violation de la propriété de *consensus* du vote en parallèle [2, 11, 12].

Ces problèmes sont évidemment cruciaux si l’on considère l’acceptabilité sociale des résultats obtenus par le vote en parallèle. D’un autre côté, le vote sur des combinaisons de valeurs (ou vote par “paquets”), qui est la seule façon d’échapper complètement à ces problèmes, est impossible en pratique, en raison de la combinatoire du problème. De fait, le vote parallèle, pour imparfait qu’il soit, est utilisé dans de nombreuses situations de référendums multiples, par exemple en Californie [5].

Les travaux dont rendent compte cet article visent à étudier la question d’un point de vue expérimental : la question est de savoir dans quelle mesure le vote parallèle peut approximer une règle de vote sur un domaine combinatoire ; autrement dit, on cherche à savoir quand, et dans quelle mesure, le vote parallèle est une bonne approximation du résultat que l’on aurait obtenu avec la méthode d’origine. Pour ce faire, nous générons aléatoirement des profils de vote combinatoires, et en particulier des profils de vote séparables ; ce problème est loin d’être trivial, vu que le ratio entre le nombre de profils séparables et le nombre de profils quelconques est très faible [7].

## 2 Background et notations

### 2.1 Vote

Soit  $\mathcal{X}$  un ensemble fini de  $m$  options (ou candidats, lorsque le vote vise à choisir des personnes). Un *vote* sur  $\mathcal{X}$  est une relation d’ordre linéaire  $\succ$  (c’est-à-dire une relation binaire irréflexive, asymétrique, transitive et totale) sur  $\mathcal{X}$ . Un *profil*  $P = (\succ_1, \dots, \succ_n)$  est une collection de  $n$  votes sur  $\mathcal{X}$ , où  $\succ_i$  est le vote du votant  $i$ . Le vote de  $i$  est censé représenter ses préférences, sous l’hypothèse que les votes sont sincères. Une *règle de vote*  $r$  est une fonction qui à chaque profil  $P$  associe une option  $r(P) \in \mathcal{X}$ .

Il existe de nombreuses règles de vote “classiques” et largement étudiées (voir par exemple [4]). Dans cet article,

<sup>1</sup>. On reviendra plus en détail en Section 2 sur deux formes de séparabilité ; il suffit ici de dire que l’argument tient pour les deux formes.

nous nous intéresserons à deux groupes de règles de vote. Tout d’abord, les règles dites de “scoring”. Étant donné un vote  $\succ$  et une option  $x \in \mathcal{X}$ , soit  $rg(\succ, x) \in \{1, \dots, m\}$  le rang de  $x$  dans  $\succ$ . On se donne maintenant un vecteur de scores  $\langle s_1, \dots, s_m \rangle$ , avec  $s_1 \geq \dots \geq s_m$  ; où  $s_i$  est le score associé au rang  $i$  : à chaque fois qu’une option est classée en position  $i$  dans un vote, la contribution de ce vote à son score total est  $s_i$ . Étant donné un profil  $P = (\succ_1, \dots, \succ_n)$ , le score de  $x$  pour  $P$  est donc  $s(P, x) = \sum_{i=1, m} s_{rg(\succ_i, x)}$ . On classe ensuite les options en fonction de leur score global  $s(P, x)$ . L’option choisie est celle qui maximise  $s(P, \cdot)$ . Entrent dans la catégorie des règles de scoring un grand nombre de règles connues, et en particulier :

- *Borda* :  $s_1 = m - 1, s_2 = m - 2, \dots, s_m = 0$ .
- *pluralité* :  $s_1 = 1, s_2 = \dots = s_m = 0$ .
- *$\frac{m}{2}$ -approbation*, en supposant pour simplifier que le nombre d’options est pair :  $s_1 = s_2 = \dots = s_{\frac{m}{2}} = 1, s_{\frac{m}{2}+1} = \dots = s_m = 0$ .

Étant donné un profil  $P$ , une option  $x$  est un *vainqueur de Condorcet* si pour tout  $y \neq x$ , une majorité de votants classe  $x$  devant  $y$ . Une règle de vote est *Condorcet-cohérente* si, pour tout profil pour lequel il existe une option vainqueur de Condorcet, c’est elle qui est choisie par la règle. Aucune règle de scoring n’est Condorcet-cohérente. Le second groupe de règles que nous avons étudié comprend des règles Condorcet-cohérentes. Pour la plupart de ces règles, le gagnant peut être déterminé à partir du *graphe de majorité* associé au profil  $P$  (qui contient un arc d’une option  $x$  vers une option  $y$  si une majorité de votants classe  $x$  devant  $y$ ), ou, plus généralement, à partir du *graphe de majorité pondéré* associé à  $P$  (qui contient, pour toute paire d’options  $x, y$ , le nombre de votants  $N_P(x, y)$  qui classent  $x$  devant  $y$ ). Les deux règles Condorcet-cohérentes que nous considérerons ici sont :

- *Copeland* : l’option gagnante est celle qui gagne le plus de duels
- *Simpson*, ou encore *maximin* : l’option gagnante est l’option  $x$  qui maximise  $\min_{y \neq x} N_P(x, y)$ .

### 2.2 Domaines combinatoires

On considère un ensemble de  $\mathcal{I} = \{A, B, C, \dots\}$  de  $p$  questions, chacune étant associée à un domaine binaire :  $D(A) = D(B) = D(C) = \dots = \{0, 1\}$ .  $\mathcal{X} = D(A) \times D(B) \times D(C) \times \dots$  est l’ensemble des *options* possibles (en anglais, “alternatives”) ou, en utilisant la terminologie du vote, des *candidats*. Le nombre d’options est donc  $m = 2^p$ . Les éléments de  $\mathcal{X}$  sont des vecteurs  $\vec{x}, \vec{x}'$  ; on pourra les écrire en concaténant les valeurs des choix. Ainsi, pour  $\mathcal{I} = \{A, B, C\}$ , l’option  $(1, 0, 1)$  désigne le choix de la valeur 1 pour la question  $A$ , de la valeur 0 pour la question  $B$  et de la valeur 1 pour la question  $C$ . On utilisera la concaténation de vecteurs de valeurs pour des séquences de questions disjointes : par exemple, pour  $\mathcal{I} = \{A, B, C, D\}$ ,  $Y = (A, B)$ ,  $Z = (C, D)$ ,  $\vec{y} = (1, 0)$ ,  $\vec{z} = (0, 1)$ ,  $\vec{y}.\vec{z}$  désigne l’option  $(1, 0, 0, 1)$ . Enfin, pour tout  $\mathcal{X} \subseteq \mathcal{I}$ ,  $D_{\mathcal{X}}$  est l’ensembles des affectations  $\vec{x}$  des

éléments de  $\mathcal{X}$  dans leurs domaines respectifs. Passons maintenant à la notion de séparabilité. Il existe deux versions différentes de cette notion. La définition la plus forte est issue de la théorie du choix multi-attributs.

**Definition 1.**  $>$  est fortement séparable si pour toute partition  $\{\mathcal{X}, \mathcal{Y}\}$  de  $\mathcal{I}$  et tous  $\vec{x}, \vec{x}' \in D_{\mathcal{X}}, \vec{y}, \vec{y}' \in D_{\mathcal{Y}}$  :

$$\vec{x}.\vec{y} > \vec{x}'.\vec{y} \iff \vec{x}.\vec{y}' > \vec{x}'.\vec{y}'$$

Toute relation de préférence issue d'un modèle de préférence par critères additifs (chaque critère correspondant à une variable de  $\mathcal{I}$ ) est fortement séparable<sup>2</sup>.

Une notion plus faible, qui est obtenue à partir de la version forte en fixant  $\mathcal{X}$  à un singleton, signifie que les préférences sur les valeurs d'une variable ne dépendent pas de la valeur des autres variables (tandis que la version forte signifie que les préférences sur les combinaisons de valeurs d'un ensemble de variables ne dépendent pas de la valeur des autres variables). Cette notion faible est équivalente à l'indépendance préférentielle mutuelle [1].

**Definition 2.**  $>$  est faiblement séparable si pour toute variable  $A \in \mathcal{I}$ , pour tous  $\vec{v}, \vec{v}' \in D_A, \vec{x}, \vec{x}' \in D_{\mathcal{I} \setminus \{A\}}$  :

$$\vec{v}.\vec{x} > \vec{v}'.\vec{x} \iff \vec{v}.\vec{x}' > \vec{v}'.\vec{x}'$$

En d'autres termes, lorsque la relation est faiblement séparable,  $v$  est préférée à  $v'$  comme valeur pour  $A$ , toutes choses égales par ailleurs (ceteris paribus). Toute relation de préférence complète et compatible avec un CP-net sans aucun arc dans le graphe de dépendance [3] respecte cette propriété de séparabilité faible.

### 3 Génération de profils séparables

Notre objectif principal est de comparer le résultat d'un vote "exact" sur un domaine combinatoire, obtenu en appliquant une règle de vote au profil complet, avec le résultat que donne le vote parallèle. On sait qu'il n'existe pas de correspondance de vote neutre qui puisse assurer que, même lorsque les préférences sont faiblement séparables, les deux résultats coïncident [2, 12]. On va donc s'intéresser à la probabilité que les deux résultats soient identiques. En raison de la grande difficulté d'une approche analytique, nous choisissons de recourir à des expérimentations sur des échantillons générés aléatoirement. L'idéal serait de pouvoir utiliser pour cette comparaison des données réelles, mais de telles données sont rarement disponibles, et à notre connaissance il n'en existe pas de taille significative sur les problèmes de référendums multiples.<sup>3</sup>

D'où la nécessité de recourir à un générateur aléatoire de profils. Le problème de la génération de profils n'est pas

2. Si les questions avaient des domaines continus, cette condition pourrait être vue comme une condition d'additivité des préférences au sens de Debreu [6].

3. Dans les données sur les référendums multiples californiens utilisées dans [5], seule la combinaison de valeurs préférée de chaque votant est disponible, mais pas la relation de préférence dans son intégralité.

nouveau ; les théoriciens du choix social utilisent souvent le modèle de la *culture impartiale*, où tous les profils sont équiprobables. Dans le cas qui nous intéresse, le nombre d'options,  $m$ , est égal à  $2^p$ ,  $p$  étant le nombre de questions. Il y a donc  $(2^p)!$  votes possibles pour un votant.

Générer un profil de vote avec une probabilité uniforme sur de tels domaines n'est pas un problème, pourvu que le nombre  $p$  de questions ne soit pas trop élevé. Cela dit, la probabilité d'obtenir un vote faiblement (et a fortiori, fortement) séparable parmi les  $2^p!$  est très faible [7]. Par exemple, pour  $p = 4$ , il existe seulement 5376 ordres fortement séparables sur  $16!$  ordres, soit un ratio de  $10^{-8}$ , et 26886144 ordres faiblement séparable, soit un ratio de  $10^{-6}$  ; le calcul exact du nombre de d'ordres faiblement ou fortement séparables est, à notre connaissance, une question ouverte pour  $p > 5$ .

La question est donc de construire un générateur de votes (faiblement ou fortement) séparables équiprobables. Une première méthode naïve mais exacte consisterait à tirer des ordres de manière équiprobable et de ne garder que ceux qui sont séparables ; vue la très faible probabilité de tirer un ordre séparable, cette méthode n'a aucun intérêt pratique. Nous présentons dans les paragraphes suivants des générateurs de préférences séparables qui, s'ils ne garantissent pas l'équiprobabilité, essaient de l'approcher autant que possible. Nous comparons ensuite leur qualité par rapport à cet objectif.

### Génération par stockage

Cette méthode consiste à construire une table contenant tous les ordres séparables normalisés : un ordre  $>$  est normalisé [7] si (i) son option préférée est  $(1, \dots, 1)$  et (ii)  $(0, 1, 1, \dots, 1) > (1, 0, 1, \dots, 1) > \dots > (1, 1, 1, \dots, 0)$ . Tout ordre séparable peut être obtenu à partir d'un ordre normalisé en permutant certaines questions et/ou en inversant certaines réponses. En ce qui concerne la séparabilité faible, cette méthode peut être envisagée pour quatre questions, mais pas plus : nos premiers tests ont montré que pour cinq questions, une telle table nécessiterait au moins 5 téraoctets (et probablement beaucoup plus). Dans le cas de la séparabilité forte, le nombre d'ordres étant plus faible, cette méthode est praticable jusqu'à six questions.

### Génération par utilité (séparabilité forte)

Cette méthode consiste à attribuer aléatoirement à chaque valeur  $j$  dans le domaine de chaque question  $i$  une utilité  $u(i, j) \in ]0, 1]$ . L'utilité additive globale de chaque option  $\vec{x}$ ,  $u(\vec{x}) = \sum_{i=1, m} u(i, x_i)$  permet ensuite de classer les options. Les ordres additifs sont fortement séparables. Kraft, Pratt et Seidenberg [9] ont montré que la réciproque est fautive dans le cas discret à partir de 5 questions. Cette méthode présente néanmoins l'intérêt d'utiliser un modèle d'utilité représentatif de bien des cas - les cas où les préférences des votants sont additives.

## Génération par réparation (séparabilité faible)

Le principe de cette méthode est de tirer un ordre  $\vec{x}_1 \succ \vec{x}_2 \succ \dots \succ \vec{x}_m$  de manière aléatoire parmi tous les ordres possibles (cet ordre a donc très peu de chances d'être séparable), puis de le réparer de manière à le rendre faiblement séparable. L'algorithme suivant effectue cette réparation :

Entrée : un ordre généré aléatoirement  $\vec{x}_1 \succ \dots \succ \vec{x}_m$

Sortie : un ordre faiblement séparable

Pour chaque question  $A \in \mathcal{I}$  faire :

  Soit  $pref[A] \leftarrow \vec{x}_1[A]$ ; % Val. préférée pour  $A$

  Soit  $W \leftarrow \mathcal{I} \setminus \{A\}$ ; % les autres questions

  Pour  $i$  allant de 2 à  $m - 1$ , si  $\vec{x}_i[A] \neq pref[A]$  faire :

    S'il existe  $i < j \leq m$  tel que  $\vec{x}_j = pref[A].\vec{x}_i[W]$   
    alors échanger  $\vec{x}_j$  et  $\vec{x}_i$ .

Pour chaque question  $A \in \mathcal{I}$  (les questions étant choisies dans un ordre aléatoire), on choisit comme valeur préférée pour  $A$  la valeur que l'option préférée,  $\vec{x}_1$ , donne à  $A$  (nous l'avons notée  $pref[A]$ ); on parcourt ensuite le reste de l'ordre en procédant à un échange chaque fois qu'on trouve deux options  $\vec{x}_i$  et  $\vec{x}_j$  telles que : 1)  $x_i$  n'a pas la valeur préférée pour  $A$ ; et 2)  $x_j$  vient après  $x_i$  et est identique à  $x_i$  sauf en ce qui concerne la valeur de la question  $A$ : cela viole la séparabilité faible de  $A$ . À la fin de ce parcours, la préférence sur chacune des questions est indépendante des autres questions. On peut montrer que rendre une variable indépendante n'affecte pas l'indépendance des variables déjà indépendantes.

Donnons un exemple avec trois variables. Supposons que le résultat du tirage aléatoire soit l'ordre  $001 \succ 100 \succ 101 \succ 000 \succ 111 \succ 110 \succ 010 \succ 011$ . Commençons par la première variable : comme  $0 \succ 1$  (en raison du fait que l'option préférée est 001, on doit d'abord intervertir 100 et 000, 111 et 011, puis 110 et 010, ce qui résulte en  $001 \succ 000 \succ 101 \succ 100 \succ 011 \succ 010 \succ 110 \succ 111$ . Considérons maintenant la deuxième variable, pour laquelle on a  $0 \succ 1$  : aucune interversion n'est nécessaire. Enfin, considérons la troisième variable, pour laquelle on a  $1 \succ 0$  : il faut intervertir 110 et 111, ce qui résulte en l'ordre faiblement séparable  $001 \succ 000 \succ 101 \succ 100 \succ 011 \succ 010 \succ 111 \succ 110$ . Il faut noter que l'ordre dans lequel les variables sont considérées peut avoir une influence sur le résultat : si l'on avait considéré les variables dans l'ordre inverse, on aurait obtenu  $001 \succ 101 \succ 000 \succ 100 \succ 011 \succ 010 \succ 111 \succ 110$ .

L'ordre ainsi réparé sera faiblement séparable, et la complexité de l'algorithme est de l'ordre de  $O(p \cdot 2^m)$  : pour chacune des  $p$  questions, on ne parcourt pas l'ordre plus de  $m$  fois,  $m$  étant le nombre (combinatoire) d'options. Cependant, ce générateur n'est pas équiprobable. En effet, la réparation peut être vue comme une recherche dans le voisinage de l'ordre initial ; alors si un ordre faiblement séparable a plus de voisins non séparables qu'un autre, il a plus de chance d'être obtenu par tirage et réparation d'un ordre aléatoire.

## Génération par exploration de treillis (séparabilité faible)

Cette méthode consiste à explorer le treillis de préférences. En effet, sous l'hypothèse que pour chaque question, la valeur 1 est préférée à la valeur 0 (puisque l'on génère un ordre normalisé), l'ordre de préférence est partiellement fixé et définit un treillis : les ordres totaux faiblement séparables sont les linéarisations de ce treillis.

Le principe de la méthode est de construire l'ordre en ajoutant les options une par une en respectant les contraintes imposées par le treillis. L'option préférée est forcément celle du haut du treillis ; il faut ensuite choisir la seconde, parmi celle qui ne sont dominées que par la première. On itère en choisissant aléatoirement à chaque itération une option qui n'est dominée que par des options déjà choisies. Chaque fois qu'il y a un choix, l'option suivante est tirée en considérant une loi uniforme entre les choix possibles. Cependant, un tirage équiprobable entre les choix possibles ne garantit pas l'équiprobabilité des ordres construits. Ceci est dû à la variation du nombre d'options accessibles selon le chemin déjà choisi. Sur la figure 3, le nombre d'options qu'il est possible de choisir après avoir choisi les 4 premières (surlignées) dépend de ces 4 premières – et donc la probabilité de l'ordre issu de ces choix varie. On peut calculer les probabilités des ordres les plus et moins fréquents de ce type de tirage. Pour  $p = 4$  elles sont de  $1/331776$  et  $1/238878720$  respectivement.

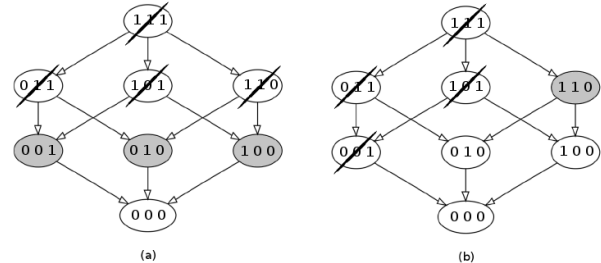


FIGURE 1 – Le treillis des options pour  $p = 3$ , après deux parcours de quatre options, (a) trois choix possibles, (b) : un choix possible.

L'idée est alors de choisir non pas la prochaine option parmi les options possibles, mais de procéder niveau par niveau (du haut du treillis vers le bas) : on choisit une option dans le niveau courant, on détermine ses emplacements possibles étant donnés les options déjà placées (de manière à respecter la propriété de séparabilité faible) : l'emplacement de cette option est alors tiré de manière équiprobable parmi ces derniers. Au sein d'un même niveau, c'est l'option qui a le moins de possibilités de placement qui est placée en premier - en effet, les probabilités de placement des options les plus restreintes ne sont pas altérées par le placement d'autres options.

Par exemple, considérons le début d'ordre suivant après l'ajout des deux premiers étages :

111  $\succ$  011  $\succ$  101  $\succ$  110. Au niveau suivant, l'option la moins restreinte est 001 puisqu'on a deux choix (et seulement un pour 100 et un pour 010). Si on insère 001 en premier, il aura une chance sur deux d'être en quatrième position, et une sur deux d'être en cinquième, sixième ou septième position (puisque 100 et 010 sont insérés après). Si l'on insère 001 en dernier, on aura bien une chance sur quatre de la retrouver à chacun des emplacements possibles.

## Comparaison des générateurs

Pour comparer la qualité des générateurs décrits ci-dessus, nous les avons testés dans le cas de 4 questions : on sait qu'il y a dans ce cas 70016 ordres normalisés faiblement séparables (et donc  $70016 \times 2^4 \times 4! = 26886144$  ordres faiblement séparables), et 14 ordres normalisés fortement séparables (et donc  $14 \times 2^4 \times 4! = 5376$  ordres fortement séparables).

Chaque générateur a été lancé au moins  $10^7$  fois pour générer un ordre normalisé, et nous avons compté le nombre d'apparitions de chaque ordre normalisé possible. Le tableau suivant donne les fréquences d'occurrence des ordres les plus et moins fréquents, le ratio entre ces deux fréquences, ainsi qu'une estimation de l'entropie de chaque générateur : l'entropie permet de mesurer à quelle point une distribution de probabilité s'approche de l'équiprobabilité, elle varie de 1 (équiprobabilité) à 0 (déterminisme).

	Stock.	Util.	Réparation	Treillis
Sép. faible	X		X	X
Sép. forte	X	X		
Fréq. max.		4/25	1/3900	1/44050
Fréq. min.		1/25	1/1400000	1/133333
fréq. max fréq.min.	1	4	360	3
Entropie	1	0,92	0,97	0,9994

Le générateur par stockage est le seul qui soit parfaitement équiprobable. Il est utilisable lorsque le nombre de questions reste faible : jusqu'à 4 pour le cas faiblement séparable et jusqu'à 6 pour le cas fortement séparable.

Comme on pouvait s'y attendre, le générateur par utilité a une mauvaise entropie lorsqu'il y a 4 questions – on sait même qu'il ne permet pas d'atteindre tous les ordres séparables à partir de 5 questions. Il a néanmoins l'avantage d'être assis sur un modèle connu et réaliste du type de préférence des votants, et c'est donc celui que nous utiliserons pour générer des préférences fortement séparables au-delà de 6 questions.

Pour générer des préférences faiblement séparables, notre expérience semble indiquer que le générateur par exploration de treillis est meilleur que celui par réparation – il l'est au moins pour 4 questions. Nous ne pouvons pas vérifier expérimentalement que ce résultat persiste au-delà de 4 questions, néanmoins nous utiliserons le générateur par exploration de treillis dans la suite.

## 4 Étude expérimentale

L'objectif de notre étude expérimentale est d'évaluer la validité du vote parallèle pour les référendums multiples. Nous considérons cinq règles de vote : Borda,  $\frac{m}{2}$ -approbation, pluralité, Simpson et Copeland. Nous étudions l'influence, sur la qualité de l'approximation, du nombre de questions, du nombre de votants, et du caractère (faiblement/fortement) séparable des préférences.

Puisque les questions sont booléennes, le vote parallèle consiste à effectuer un vote majoritaire sur chaque question. Pour les profils non séparables, on effectue ce vote de manière optimiste (ce qui est justifié par l'observation de comportements réels) : les votants choisissent les valeurs correspondant à leur option préférée. Afin de limiter les *ex aequo*, on choisira un nombre de votants  $n$  impair. Le résultat du vote parallèle (l'option choisie) est alors comparé au résultat du vote sur le domaine combinatoire effectué selon la règle de vote  $r$  choisie. On comptabilise le nombre de succès (résultats identiques avec les deux procédures). S'il y a des *ex aequo* lors du vote combinatoire, et que le résultat du vote parallèle figure parmi ces *ex aequo*, le test est considéré comme un succès. Pour chacune des expériences ci-dessous, on reproduit le test 10000 fois (chaque point correspond à une moyenne sur 10000 profils).

Nous examinons tout d'abord le cas  $p = 4$ , pour lequel nous avons des générateurs parfaitement équiprobables (les générateurs par stockage pour la séparabilité faible et la séparabilité forte, un simple tirage aléatoire pour le cas général). Les résultats sont présentés sur la figure 2.

On observe immédiatement que pour les trois règles de scoring, le taux de réussite passe en dessous de 60% à partir de 7 votants : dans 6 cas sur 10 au moins, la règle parallèle donne comme gagnante une option différente de celle qu'aurait choisi le vote combinatoire. Ce taux est particulièrement mauvais pour la pluralité et la  $m/2$  approbation. Nous analysons maintenant l'influence sur le taux de réussite de l'approximation de la séparabilité du profil, du nombre de questions et du nombre de votants.

### 4.1 Influence de la séparabilité

Sans surprise, nos résultats sont cohérents avec les prévisions de travaux de [10] et [8] : la séparabilité augmente le taux de réussite. Sauf pour la pluralité, les taux de réussite sont bien meilleurs dans les cas séparables que dans le cas général. Ces résultats se confirment lorsque l'on augmente de nombre de votants (cf. figure 2) ou le nombre de questions (voir figure 3 pour Borda, les autres règles, pluralité excepté, se comportant de façon similaire). Notons que le taux de réussite semble légèrement meilleur dans le cas faiblement séparable que dans le cas fortement séparable. L'explication de cette légère différence reste une question ouverte.

Dans le cas de la pluralité, nous obtenons les mêmes résultats que les profils soient séparables ou non, ce qui est naturel, puisque la pluralité, comme le vote parallèle, ne s'intéresse qu'au premier choix de chacun des votants : comme

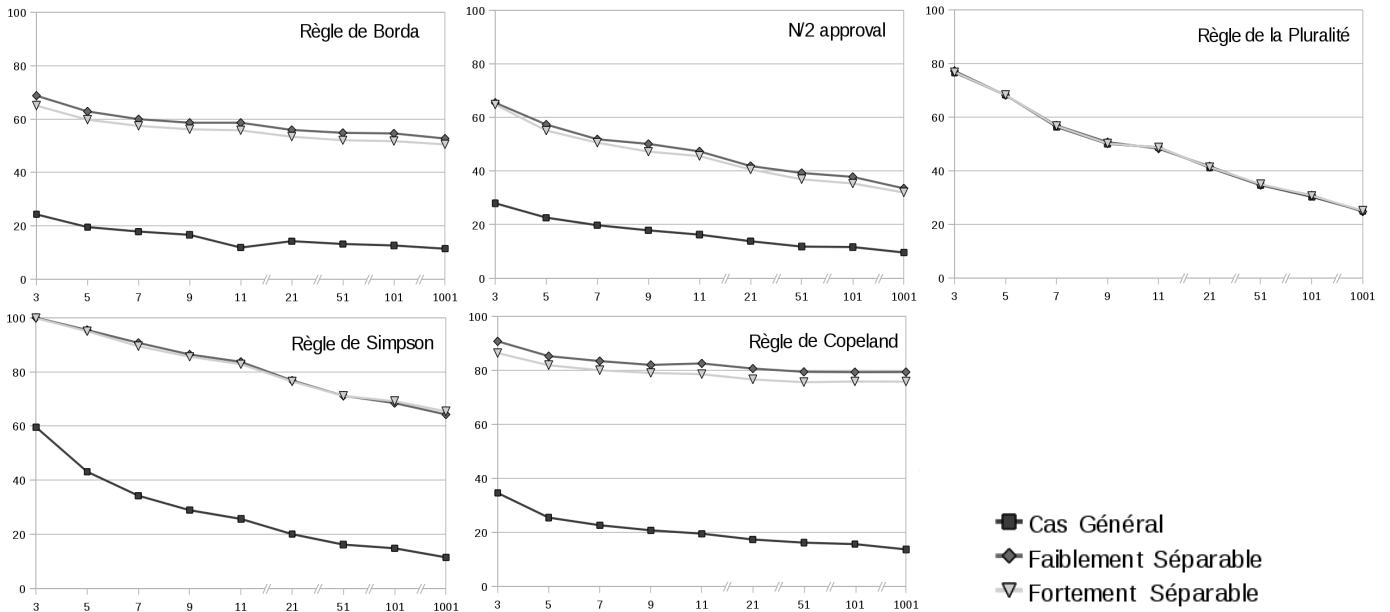


FIGURE 2 – Taux de réussite du vote parallèle en fonction du nombre de votants, 4 questions

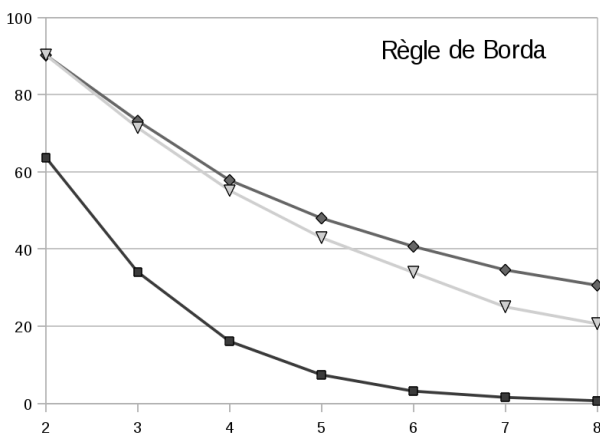


FIGURE 3 – Règle de Borda : taux de réussite du vote parallèle en fonction du nombre de questions, 11 votants

on ne tient compte que de la première option de l'ordre de préférence, la séparabilité n'a aucune influence.

Pour la suite des expériences, nous nous intéresserons uniquement aux expérimentations menées sur des profils faiblement ou fortement séparables, et ceci pour deux raisons : (a) dans le cas général de préférences sans hypothèse de séparabilité, l'approximation est médiocre ; (b) en l'absence d'hypothèse de séparabilité, il y a pas une façon unique, pour les votants, de reporter leurs préférences sur chacune des questions (même si l'hypothèse d'un comportement optimiste, que l'on a choisie, est souvent raisonnable) ; les résultats sont donc à prendre avec plus de circonspection que dans le cas des préférences séparables.

## 4.2 Comportement des règles

Comme on l'a dit plus haut, on ne génère maintenant que des profils (faiblement ou fortement) séparables.

La figure 4 synthétise, pour le cas  $p = 4$ , les taux de réussite des règles en fonction du nombre de votants pour le cas faiblement séparable (le cas fortement séparable donnerait des résultats similaires) : le taux de réussite décroît quand le nombre de votants augmente.

Il dépend, et de manière plus forte encore, du nombre de questions : dans une seconde série de tests dont rendent compte la figure 5, nous avons mesuré ce taux en fonction du nombre de questions pour 7 votants, pour les règles qui donnaient des résultats satisfaisants pour le premier test, à savoir Borda, Copeland et Simpson. Ces tests ont été réalisés pour le cas de préférences faiblement séparables, et donc à partir du générateur par exploration de treillis ; les résultats pour le cas fortement séparables sont similaires.

Le taux de réussite décroît donc en fonction du nombre de votants et surtout en fonction du nombre de questions. Encore une fois, il est très rapidement très mauvais pour la règle de la pluralité, et pour celle du  $\frac{m}{2}$ -approbation. Il est moins mauvais pour Borda, mais se stabilise quand même autour des 50% dès 5 questions (7 votants) ou 10 votants (4 questions). Enfin, il est bien meilleur pour les deux règles Condorcet-cohérentes, surtout Copeland. En analysant les résultats plus en détails, on s'aperçoit que les cas de divergence entre les règles parallèles et combinatoires sont des cas où il n'y a pas de vainqueur de Condorcet.

## 4.3 Qualité de l'approximation

À partir de 5 questions, le taux de réussite de l'approximation chute en dessous des 50% pour toutes les règles de scoring (en dessous de 80% pour les règles de Condorcet).

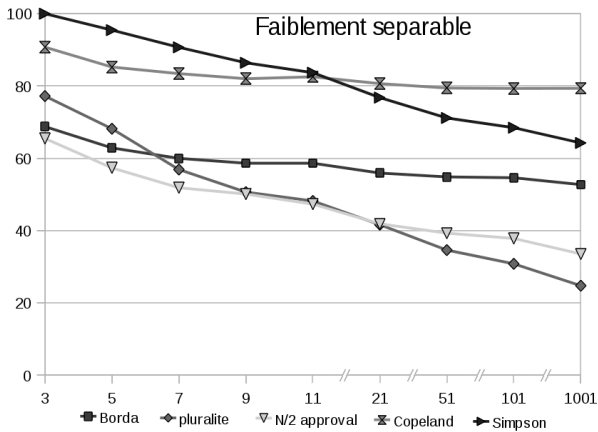


FIGURE 4 – Variation du taux de réussite en fonction du nombre de votants, ordres faiblement séparables, 4 questions

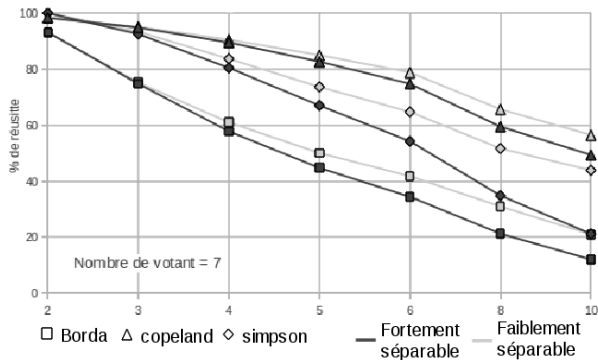


FIGURE 5 – Variation du taux de réussite en fonction du nombre de questions, 7 votants

cet Cohérentes) : le résultat du référendum est une fois sur deux différent de celui qu'une application rigoureuse d'une règle de scoring aurait donné, ce qui est en soit déjà problématique. Cela dit, on peut se demander quelle est, pour les votants, "l'ampleur" de l'erreur du vote parallèle - elle peut s'avérer finalement assez faible. On peut d'abord mesurer, pour chaque règle, l'écart entre le score accordé par la règle entre l'option choisie par le vote combinatoire et l'option choisie par le vote parallèle. Les tableaux suivants donnent ces écarts pour 7 votants ; les trois premières lignes fournissent les scores moyens dans le cas général (l'échantillon contient des cas de succès comme des cas d'échec) ; les lignes suivantes approfondissent les résultats pour les sous groupes de cas où la règle parallèle est en échec.

On observe que ce sont les règles qui sont le mieux approximées en termes de taux de réussite (Borda, Copeland et Simpson) qui se dégradent le moins quand l'option choisie par le vote parallèle n'est pas l'option optimale : il y a moins de 3% d'écart de score pour ces règles. Les résultats sont bien plus mauvais pour la pluralité et la  $\frac{m}{2}$ -approbation.

Une autre approche, indépendante de la règle testée,

	Borda	Plur.	N/2	Cope.	Simp.
<b>Cas général</b>					
score moyen du vainqueur réel	72,71	1,9	5,73	14,22	3,47
score moyen du vainqueur en //	70,98	1,3	5,08	13,94	3,38
taux d'échec (%)	39,18	42,97	47,26	16,21	9,26
<b>Cas d'échec</b>					
score moyen du vainqueur réel	70,59	1,99	5,77	13,36	3
score moyen du vainqueur en //	66,18	0,6	4,39	11,65	1,97
distance moyenne	4,41	1,39	1,38	1,71	1,02
distance max	21	3	4	8	2
distance min	1	1	1	1	1

FIGURE 6 – écart de score, 7 votants et 4 questions, profils faiblement séparables.

consiste à mesurer la satisfaction des votants pris individuellement lorsque les résultats des deux votes diffèrent. Elle sous-entend une notion sous-jacente d'intérêt pour les votants - bref, un modèle par utilité. Les courbes suivantes donnent pour chaque règle le taux de votants qui peuvent s'estimer lésés par le vote parallèle lorsque cette règle ne fournit pas l'option qu'aurait élue le vote combinatoire ( $p = 4$ ). Par exemple, pour 7 votants et la règle de Borda, le taux de réussite était de l'ordre de 60%. Dans les 40% de cas restants, il apparaît que près d'un votant sur deux est finalement content : son utilité est meilleure que celle qu'il aurait obtenue avec la règle exacte. En fait, les courbes montrent qu'en cas d'échec de l'approximation, le taux de votants finalement satisfaits en termes d'utilité tend vers un sur deux, c'est-à-dire la satisfaction que le votant aurait obtenue par un tirage aléatoire du gagnant ; les approximations sont donc assez discutables en termes d'utilité.

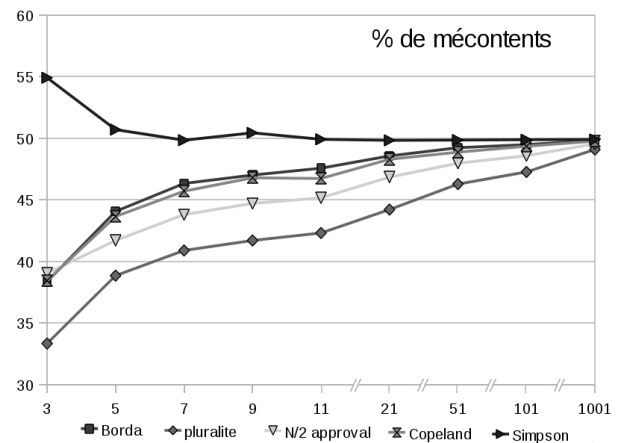


FIGURE 7 – Taux de votants mécontents de la règle parallèle quand elle est en échec



## 5 Conclusion

Dans cette étude préliminaire, nous avons tout d'abord étudié la question de la génération aléatoire équiprobable de profils de vote séparables. Nous avons proposé et testé plusieurs pistes, retenant finalement une méthode de génération par exploration de treillis, pour le cas de la séparabilité faible et une méthode fondée sur un modèle d'utilité, pour la séparabilité forte. La première obtient de bon résultats en termes d'entropie, la seconde à l'avantage d'être fondée sur un modèle connu et axiomatisé de la structure préférences des votants.

Pour conclure sur la validité pratique du vote parallèle, nos résultats sont assez négatifs. Ils soulignent que le vote parallèle est une mauvaise approche pour les règles de scoring, et en particulier pour la pluralité : pour trois règles de scoring étudiées, le gagnant du vote parallèle est différent du gagnant qu'aurait donné le vote combinatoire, dans plus de 70% des cas, dès 4 questions et 5 votants. Ces résultats se dégradent encore si l'on augmente le nombre de questions et le nombre de votants. Les résultats sont moins mauvais pour les deux règles Condorcet-cohérentes étudiées, et en particulier Copeland, qui reste à plus de 80% de réussite pour  $p = 4$  (ce qui reste une hypothèse raisonnable pour les référendums multiples réels).

Cette étude a permis de vérifier les résultats prédits par la théorie, à savoir que les résultats sont bien meilleurs dans le cas fortement séparable que dans le cas général. Cette étude étudie de plus l'influence de la propriété, moins exigeante, de séparabilité faible : les résultats semblent indiquer que la séparabilité faible suffit pour que l'approximation donne des résultats convenables, dans la mesure du possible.

Cela dit, cette étude s'est focalisée sur des critères assez immédiats, comme le taux de réussite de la règle ou la satisfaction des votants. Il faudrait la compléter par une étude plus fine de la dégradation en cas d'échec, en testant en particulier combien de fois les approximations parallèles donnent des résultats paradoxaux (par exemple, une option Pareto-dominée). Il faudrait de plus l'affiner et la compléter en testant la robustesse des règles en fonction d'un paramètre plus fin sur la séparabilité, comme le pourcentage de votes séparables dans les profils.

Enfin, puisque le vote parallèle s'avère être une mauvaise solution, il faut se poser la question de l'existence d'autres techniques de vote ou de représentation des préférences, qui permettraient une bonne approximation de la règle combinatoire.

## Références

- [1] F. Bacchus and A. J. Grove. Graphical models for preference and utility. In *UAI*, pages 3–10, 1995.
- [2] J.-P. Benoit and L.A. Kornhauser. Only a dictatorship is efficient or neutral. Technical report, NYU Law School, 2006.
- [3] C. Boutilier, R. I. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets : a tool for representing

and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21 :135–191, 2004.

- [4] S. Brams and P. Fishburn. Voting procedures. In K. Arrow, A. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare, Volume 1*, pages 173–236. Elsevier, 2002.
- [5] S. Brams, W. Zwicker, and M. Kilgour. The paradox of multiple elections. *Social Choice and Welfare*, 15(2) :211–236, 1998.
- [6] G. Debreu. Topological methods in cardinal utility theory. In K.J. Arrow, S. Karlin, and P. Suppes, editors, *Mathematical Methods in the Social Sciences*, pages 16–26. Stanford University Press, 1960.
- [7] J. Hodge. Permutations of separable preference orders. *Discrete Appl. Math.*, 154 :1478–1499, June 2006.
- [8] J. Hodge and P. Schwallier. How does separability affect the desirability of referendum election outcomes ? *Theory and Decision*, 61(3) :251–276, November 2006.
- [9] C. H. Kraft, J. W. Pratt, and A. Seidenberg. Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, 30(2) :408–419, 1959.
- [10] D. Lacy and E. M. S. Niou. A problem with referenda. *J. of Theoretical Politics*, 12(1) :5–31, 2000.
- [11] İ. Özkal-Sanver and M.R. Sanver. Ensuring Pareto optimality by referendum voting. *Social Choice and Welfare*, 27(1) :211–219, August 2006.
- [12] L. Xia and J. Lang. A dichotomy theorem on the existence of efficient or neutral sequential voting correspondences. In *Proceedings of IJCAI-09*, pages 342–347.