# Que sait-on du résultat d'une élection avant que tous les votants se soient exprimés?

Guillaume Ravilly-Abadie, Yann Chevaleyre, Mathieu Lacroix, Jérôme Lang, Nicolas Maudet

LAMSADE, CNRS & Université Paris-Dauphine

**Résumé** : Dans cet article, nous nous intéressons à des situations de votes où seulement un profil partiel est connu, partiel dans le sens où nous ne connaissons les votes que d'un sous-ensemble des votants. Dans ces situations, on peut distinguer les candidats étant des *gagnants possibles* (c'est-à- dire ceux qui ont encore une chance d'être le vainqueur une fois les derniers votes connus) de ceux qui ne le sont pas. Notre but dans ce travail est d'explorer comment le degré de complétion du profil de vote affecte le nombre de ces gagnants possibles. Cette information est pertinente à divers titres : elle permet de donner un résultat partiel dans le cas où le processus de vote doit être stoppé ; elle permet de guider le processus d'élicitation des préférences des votants restants ; et, de façon plus fondamentale, elle permet de comparer les différentes procédures de votes en fonction du degré de détermination du résultat, à degré de complétion égal du profil de vote. Nous fournissons des résultats expérimentaux montrant l'influence que peut avoir le degré de complétion pour différentes règles de vote.

## 1 Introduction

Le vote est l'un des moyens les plus utilisés pour prendre des décisions communes. On applique habituellement une règle de vote à un ensemble de votes complet (appelé un *profil*). Or, il existe de nombreux contextes où les votes ne sont pas obtenus tous au même moment. Par exemple, lors des élections politiques dans certains pays, les votes des électeurs résidant à l'étranger ne sont connus que seulement quelques jours après les autres. Ou encore, lorsqu'il s'agit de décider de la date d'une réunion, il arrive souvent que certains participants expriment leurs préférences plus tard que les autres. Dans de telles situations, on peut vouloir prétraiter l'information donnée par le sous-électorat constitué par les votants qui se sont déjà exprimés afin de préparer le terrain pour le moment où les retardataires auront exprimé leurs votes. Cette piste de recherche est explorée dans Chevaleyre *et al.* (2009), où l'on détermine, pour quelques règles de vote bien connues, la quantité d'information nécessaire pour compiler un ensemble partiel de votes. On peut vouloir aussi s'intéresser à l'ensemble des candidats qui "restent en lice", c'est-à-dire qui ont encore une chance de gagner, après qu'une partie des votants s'est exprimée – de tels candidats sont appelés *gagnants possibles*. La question générale qui nous intéresse ici est la suivante : étant donnés $p$ candidats et $n$ votants parmi lesquels seulement $k$ se sont exprimés, que peut-on dire du nombre de gagnants possibles ? En particulier, si l'on se donne une distribution de probabilité sur les profils de vote possibles, on aimerait connaître la loi de probabilité du nombre de gagnants possibles, en fonction de $p$, $n$, $k$ et de la règle utilisée ; ce qui nous donnera, *a fortiori*, les trois informations suivantes : (a) le nombre minimal de gagnants possibles, (b) leur nombre moyen, et enfin, (c) la probabilité qu'il en reste au moins deux, c'est-à -dire que le résultat du vote ne soit pas entièrement déterminé.

La question (a) ci-dessus a été traitée (au moins en partie) dans la littérature : ainsi, Moulin (1988)[p.272] identifie, pour quelques règles de vote, le nombre de candidats qu'une coalition de $k$ votants peut bloquer. Les questions (b) et (c) n'ont, à notre connaissance, pas été traitées. Or, ces questions sont importantes, pour plusieurs raisons.

D'abord, étudier le nombre de candidats qu'une coalition de votants peut bloquer en moyenne revient à étudier de manière quantitative le pouvoir de manipulation d'une règle de vote par une coalition. En effet, plus une coalition d'un nombre donné de votants peut, en moyenne, bloquer de candidats, plus la règle de vote est, dans un certain sens, manipulable. Deux indicateurs particulièrement pertinents sont la taille de la plus petite coalition permettant de faire élire un candidat (manipulation *constructive*) et la taille de la plus petite coalition permettant de bloquer un candidat (manipulation *destructive*). Le théorème de Gibbard et Satterwaite dit que dès qu'il y a au moins 3 candidats, toute règle de vote satisfaisant la non-imposition (tout candidat est le vainqueur pour au moins un profil) est manipulable, c'est-à -dire qu'il existe des situations où un votant a intérêt à exprimer des préférences non sincères. Des versions quantitatives de ce théorème ont été établies récemment par Friedgut *et al.* (2008); Xia & Conitzer (2008b), qui montrent que toute règle de vote satisfaisant un certain ensemble de conditions est manipulable par une proportion non négligeable des profils. Ces résultats restent théoriques, et l'étude empirique de la distribution du nombre de gagnants possibles permettra de voir d'un oeil différent la manipulabilité de diverses règles de vote.

Il existe par ailleurs des travaux sur la taille moyenne des coalitions permettant de manipuler une élection (Slinko (2006); Pritchard & Slinko (2006)), mais la notion de manipulation est ici assez éloignée de la notion de gagnant possible : elle suppose le profil $P$ entièrement connu, et une manipulation par une coalition est la donnée, pour chacun des membres de la coalition, d'un vote *différent* de son vote tel qu'il est dans $P$ (donc non sincère), permettant de faire élire un autre candidat que le vainqueur pour $P$, qui soit préféré à ce dernier par tous les membres de la coalition. A l'inverse, dans les travaux de choix social computationnel, en particulier Conitzer & Sandholm (2002a); Zuckerman *et al.* (2009); Xia *et al.* (2009), une manipulation constructive par une coalition est la donnée d'un vote, pour chacun des membres de la coalition, permettant de faire élire un candidat donné $x$ ; le lien avec les gagnants possibles est alors évident : $x$ est un gagnant possible s'il existe une manipulation constructive, par la coalition des votants restants, faisant gagner $x$.

Ensuite, la distribution des gagnants possibles permet d'apprécier la difficulté du processus d'élicitation permettant de déterminer le résulat d'un vote (Conitzer & Sandholm (2002b)). Dans le cas où le résultat n'est pas complètement déterminé, il est intéressant d'avoir une estimation du nombre de votants dont il faudra éliciter les préférences pour déterminer le résultat. Par ailleurs, même lorsqu'il faudra, dans de nombreux cas, tenir compte des candidats qui ne sont pas des gagnants possibles lorsque l'on élicite les préférences des votants restants, la connaissance des gagnants possibles permet de focaliser la discussion et la négociation sur ces derniers, et permet de guider le processus d'élicitation.

Enfin, ces indicateurs relatifs au nombre de gagnants possibles sont tout simplement une mesure du caractère décisif d'une règle ou d'une correspondance de vote, et c'est une propriété des règles de vote qu'il est pertinent d'étudier pour elle-même.

Les résultats obtenus dans cet article suivent la méthodologie suivante : on fixe une règle de vote ; on fait l'hypothèse de la culture impartiale (tous les profils ont la même probabilité de survenir) ; puis pour certaines valeurs de $p$, $n$ et $k$, on estime expérimentalement, au moyen d'un algorithme de Monte Carlo, la distribution de probabilité du nombre de gagnants possibles. Cet algorithme nécessite de pouvoir calculer, pour chaque profil généré aléatoirement, l'ensemble des gagnants possibles, ce qui est souvent un problème difficile Xia & Conitzer (2008a). On devra donc, pour certaines des règles de vote étudiées, construire (ou parfois, réutiliser) des algorithmes aussi efficaces que possible pour le calcul des gagnants possibles.

La section 2 donne un rappel nécessaire concernant les règles de vote et leur application à des profils incomplets. La Section 3.1 expose la méthodologie générale de notre approche. Dans chacune des sections suivantes on examine une règle de vote, ou une famille de règles de vote, particulière : la pluralité, le véto et plus généralement les règles de $k$-approbation en section 3.2, Pluralité à deux tours en section 3.3, Borda en section 3.4, et enfin, Copeland en section 3.5. Finalement, en section 3.6 on fait une étude comparative des résultats obtenus pour différentes règles de vote.

# 2 Préliminaires

## 2.1 Règles de vote

Dans la suite de l'article, on considère un ensemble $X = \{x_1, \ldots, x_p\}$ de $p$ candidats, et un ensemble $N$ de $n$ votants (qui constituent la totalité des votants).

Un *vote* est un ordre linéaire sur $X$.On note en général les votes de la faaçon suivante : $a \succ b \succ c$ est noté *abc*, etc. Pour tout $m \geq 0$, un $m$-*profil* sur $X$ est un $m$-uplet $P = \langle V_1, \ldots, V_m \rangle$ de votes. Lorsque $m < n$ (resp. $m = n$), un tel profil est *partiel* (resp. *complet*). Soit $\mathcal{P}_X^m$ l'ensemble de tous les $m$-profils sur $X$. On désigne par $\mathcal{P}_X^*$ l'ensemble de tous les $m$-profils pour $m \geq 0$, *c.a.d.*, $\mathcal{P}_X^* = \cup_{k \geq 0} \mathcal{P}_X^k$.

Une *règle de vote* est une fonction $r$ de $\mathcal{P}_X^n$ dans $X$. Comme la définition habituelle des règles de vote n'exclut pas la possibilité de gagnants ex-aequo, on suppose que les ex-aequo sont départagés par un ordre sur les candidats fixé a priori.

Les différentes règles de vote dont il est question dans cet article seront présentées dans la section correspondante.

## 2.2 Gagnants possibles et nécessaires

Sur la base de profils incomplets, il est déjà possible de distinguer des candidats : certains sont des gagnants possibles (Konczak & Lang (2005)), au sens où ils peuvent encore gagner l'élection pour au moins une complétion du profil. Formellement : étant donné un profil partiel $P = \langle V_1, \ldots, V_m \rangle$ et une règle de vote $r$, un candidat $x \in X$ est un *gagnant possible* pour $r$ et $P$ s'il existe un $(n-m)$-profil $Q = \langle V_{m+1}, \ldots, V_n \rangle$ tel que $r(P \cup Q) = x$. On désigne par $Poss_r(P)$ l'ensemble des candidats possibles pour $r$ et $P$. Un candidat $x \in X$ est un *gagnant nécessaire* pour $r$ et $P$ si pour tout $n-m$-profil $Q = \langle V_{m+1}, \ldots, V_n \rangle$ on a $r(P \cup Q) = x$. On désigne par $Nec_r(P)$ l'ensemble des candidats nécessaires pour $r$ et $P$. Bien entendu, on a $Nec_r(P) \subseteq Poss_r(P)$, $Poss_r(P) \neq \emptyset$, et on remarquera aussi que puisque les règles de vote sont déterministes, $Nec_r(P)$ contient au plus un candidat et que $Nec_r(P) \neq \emptyset$ si et seulement si $Poss_r(P)$ est un singleton.

Le propos de cet article est d'étudier la manière dont la complétion progressive du profil va influencer l'ensemble des gagnants possibles.

## 2.3 Fonctions de veto anonymes

On peut s'intéresser en premier lieu à deux seuils importants : le premier est la plus petite valeur de $m$ permettant d'éliminer un candidat de l'ensemble des gagnants possibles ; le second est la plus petite valeur de $m$ permettant à une coalition d'imposer son gagnant.

$$A(r,p,n) = min\{m | \exists Q \in \mathcal{P}_X^m \text{ tq. } |Poss_r(Q)| \neq p\}$$
$$B(r,p,n) = min\{m | \exists Q \in \mathcal{P}_X^m \text{ tq. } |Poss_r(Q)| = 1\}$$

Il est possible de calculer asymptotiquement les valeurs de $A$ et de $B$ pour la classe générale des *règle de scoring*, comme le montre la proposition suivante dont la preuve est omise.

**Proposition 1**
*Soit $\vec{s} = \langle s_1, \ldots, s_p \rangle$ un vecteur tel que $s_1 \geq \ldots \geq s_p$ et $s_1 = 1$, $s_p = 0$. Soit $r_{\vec{s}}$, la règle de scoring basée[1] sur $\vec{s}$. Lorsque $n$ croit vers l'infini, on a*

$$[A(r,p,n), B(r,p,n)] \to \left[ \frac{n}{2}, \frac{n}{2 - \frac{\sum_{j=2}^{p-1} \alpha_j}{p-1}} \right]$$

Ainsi, pour la règle de pluralité, $B(r,p,n) \to \frac{n}{2}$ et pour la règle de veto, $B(r,p,n) \to n\frac{p-1}{p}$.

Notons que cette notion peut être généralisée. Moulin définit dans Moulin (1988)[p.272] la notion de *fonction de veto anonyme (fva)*. Une fonction de veto anonyme $v$ est une fonction qui, pour un nombre de votants $m$, renvoie le nombre maximal de candidats que n'importe quelle coalition de cette taille peut bloquer. Le lien avec nos notions est

[1] la contrainte sur $s_1$ et $s_p$ n'implique pas de perte de généralité

direct : le nombre de gagnants possibles est $p - v(m)$, le nombre de candidats non bloqués. Selon cette notation, $A$ est donc la plus petite valeur telle que $v(m) > 0$, tandis que $B$ est la plus petite valeur telle que $v(m) = p - 1$. Moulin observe dans Moulin (1988)[Example 10.2] que pour les règles Condorcet-consistantes, $v(m) = 0$ quand $m \leq \frac{1}{2}n$, tandis que $v(m) = p - 1$ quand $m > \frac{1}{2}n$. Moulin caractérise également la *vfa* de Borda. On a également $v(m) = 0$ quand $t \leq \frac{1}{2}n$, mais pour cette règle, il faut au minimum deux tiers des votes pour imposer un vainqueur ($v(m) = p - 1$ quand $m > \frac{2}{3}n$).

Pourtant cette notion ne dit pas tout : bien sûr, il est théoriquement possible qu'une élection ait déjà déterminé son vainqueur juste après qu'une majorité de personnes s'est exprimée. En ce sens, les valeurs données par les *vfa* constituent la borne inf. sur le nombre de gagnants encore possibles. Mais c'est une borne théorique seulement. En pratique, il faudra en général bien plus de votes pour que la probabilité d'avoir un résultat déterminé doit non négligeable. (Notons qu'il n'est pas pertinent de se poser la question d'une borne sup, puisqu'il est toujours possible de construire des profils pathologiques laissant leur chance à tous les candidats jusqu'à l'ultime vote, ou presque). Mais ces notions ne permettent par exemple en rien de distinguer le comportement des règles Condorcet-consistantes. Dans la suite de cet article on cherche à déterminer cette valeur en moyenne.

# 3 Calcul expérimental du nombre de gagnants possibles

On s'intéresse à présent à la question du calcul du nombre moyen de gagnants possibles pour les différentes règles de vote, en fonction du nombre de candidats restants. Les résultats de la section précédente donnent une première indication sur le comportement des règles, on souhaite poursuivre nos investigations dans ce sens.

## 3.1 Méthodologie

Toute approche expérimentale dans le domaine du choix social pose la question épineuse de la génération aléatoire des profils qui seront utilisés lors des tests. On rencontre généralement deux approches (voir par exemple (Berg & Lepelley (1992)) :

- l'hypothèse de la *culture neutre*, selon laquelle tous les profils ont la même probabilité de survenir ;
- l'hypothèse de la *culture neutre et anonyme*, qui suppose également une distribution uniforme de probabilités, mais sur les "situations de vote" [2].

Dans cet article, nous adoptons l'hypothèse de la *culture neutre*. Plus précisément, à partir de la distribution aléatoire et uniforme sur $\mathcal{P}_X^1$ que nous nommerons $\mathcal{D}$, nous calculerons une valeur approchée de l'espérance du nombre de gagnants possibles :

$$\mathbb{E}_{P \sim \mathcal{D}^m} \left[ |Poss_r(P)| \right] = \mathbb{E}_{V_1 \ldots V_m \sim \mathcal{D}} \left[ \left| \left\{ r(V_1 \ldots V_n) : V_{m+1} \ldots V_n \in \mathcal{P}_X^{n-m} \right\} \right| \right]$$

[2]Une situation de vote est un vecteur enregistrant pour chacun des $p!$ ordres de préférence possible le nombre de votants ayant cet ordre de préférence, ce qui est suffisant pour les règles anonymes, voir Berg & Lepelley (1992), et aussi Chevaleyre *et al.* (2009) pour une généralisation de cette idée.

Evidemment, pour chaque profil généré $P$, pour chaque candidat $x$, nous devons répondre à la question de savoir s'il est ou non un gagnant possible (donc si $x \in Poss_r(P)$). Au vu du nombre de tests envisagés, on comprend qu'il est crucial que cette question puisse trouver réponse rapidement. A cet égard, les règles de vote sont plus ou moins dociles. La difficulté algorithmique de la question : "ce candidat est-il possible ?" correspond au problème de savoir si une coalition de taille donnée peut manipuler une élection (Zuckerman *et al.* (2009)). En effet, notre problème peut être reformulé ainsi : la "coalition" des votants restants est-elle en mesure de faire gagner un candidat donné ?

A l'aide d'un algorithme de Monte-Carlo nous estimons la distribution du nombre de gagnants possibles pour chaque règle de vote. Le nombre de tests est de $10000$ sauf pour Copeland (où seulement $1000$ tests ont été effectués).

Dans les sections qui viennent nous présentons les résultats pour les différentes règles de vote. Dans certains cas (pluralité, véto, $k$-approbation), des algorithmes polynomiaux existent. Dans d'autres cas, des algorithmes approchés existent (voir notamment Zuckerman *et al.* (2009), pour Borda) ; nous réutilisons ces algorithmes, mais nous avons cependant besoin d'aller plus loin, puisqu'il nous faut calculer le nombre exact de gagnants possibles. Dans d'autres cas (Copeland, Simpson), aucun algorithme n'existe, et l'une des contributions de cet article est de proposer de tels algorithmes.

## 3.2 Pluralité, véto et $k$-approbation

Etant donnés $P \in \mathcal{P}_X^m$ et $x \in X$, soit $n(P,i,x)$ le nombre de votes dans $P$ qui classent $x$ en position $i$, and $ntop(P,x) = n(P,1,x)$ le nombre de votes dans $P$ qui classent $x$ en première position. Soit $\vec{s} = \langle s_1,\ldots,s_p\rangle$ un vecteur tel que $s_1 \geq \ldots \geq s_p$. La *règle de scoring* induite par $\vec{s}$ est définie par : $r_{\vec{s}}(P)$ est le candidat maximisant $S_{\vec{s}}^P(x) = \sum_{i=1}^m s_i.n(P,i,x)$. La règle de *pluralité* est la règle de scoring correspondant au vecteur $\langle 1,0,\ldots,0\rangle$, et la règle de *veto* au vecteur $\langle 1,1,\ldots,1,0\rangle$. Enfin, pour $1 \leq k \leq p-1$, la règle de $k$-approbation correspond au vecteur $\langle 1,\ldots,1,0,\ldots,0\rangle$ (où les $k$ premières composantes sont égales à 1). Notons que l'on retrouve la pluralité lorsque $k=1$ et le véto lorsque $k=p-1$. Notons $r_k$ la règle de $k$-approbation.

Le calcul des gagnants possibles pour les règles de $k$-approbation ne pose aucune difficulté, et peut être fait en temps polynomial. En effet, la Proposition 3.5 dans Meir *et al.* (2008) montre que les gagnants possibles pour la règle de $k$-approbation peuvent être calculé en exploitant la propriété suivante :

Soit $P \in \mathcal{P}_X^m$ un profil partiel quelconque et $x$ un candidat. Définissons inductivement le profil partiel $P_i$. D'abord, $P_0 = P$, puis $P_{i+1}$ est égal au profil $P_i$ auquel un vote aura été rajouté. Ce vote est construit de telle sorte que le candidat $x$ soit approuvé ainsi que les $k-1$ candidats ayant obtenu le plus petit score sur $P_i$. Alors, $x \in Poss_{r_k}(P)$ si et seulement si $x = r_k(P_{n-m})$. Par conséquent, pour calculer les gagnants possibles sur $P$, il suffit de construire $P_{n-m}$ pour chaque candidat $x$ et de vérifier si $x$ est le gagnant du profil.

Par exemple, soit $X = \{a,b,c,d\}, m=5, n=8, P = \langle abcd, abcd, bcad, bcad, cdab\rangle$ et $x=a$. Prenons $k=3$ (correspondant donc au véto), avec la priorité $a > b > c > d$ pour départager les ex-aequo. Les scores partiels sont les suivants : $a \mapsto 2, b \mapsto 4, c \mapsto$

$3, d \mapsto 1$. Le sixième votant approuve alors $a$, $c$ et $d$ (ou, de manière équivalente, exprime un véto pour $b$), et les nouveaux scores partiels sont $a \mapsto 3, b \mapsto 4, c \mapsto 4, d \mapsto 2$ ; le votant suivant approuve $a$, $b$ et $d$, et les nouveaux scores partiels sont $a \mapsto 4, b \mapsto 5, c \mapsto 4, d \mapsto 3$ ; le dernier votant approuve alors $a$, $c$ et $d$, et les scores finaux sont $a \mapsto 5, b \mapsto 5, c \mapsto 5, d \mapsto 4$. $a$ est le gagnant pour le profil complet, donc il est gagnant possible pour $P$. On remarquera que c'est aussi le cas pour $b$ et $c$, mais pas pour $d$.

Remarquons que l'algorithme précédent fonctionne en $O(p.(m-n))$. Calculer les gagnants possibles pour $P$, qui nécessite de considérer les candidats les uns après les autres, a donc une complexité en $O(p^2.(m-n))$.

A présent que l'on a un algorithme polynomial pour calculer les gagnants possibles pour les règles de $k$-approbation, on peut réaliser les expérimentations, comme on l'a décrit en Section 3.1. Nous obtenons les résultats suivants.

On constate sur la figure 1 que plus $k$ est grand, plus le résultat du vote est proche d'être déterminé, avec à une extrémité la règle de veto (pour laquelle le nombre de gagnants possibles est minimal jusqu'à des valeurs de $K(= m)$ très élevées) et la pluralité à l'extrémité opposés. Cette observation ne vaut cependant plus pour des valeurs de $K$ très élevées.

Les figures 2 et 3 montrent l'évolution de la distribution de probabilité sur le nombre de gagnants possibles lorsque le nombre $m$ de votes déjà connus évolue. A première vue, elles montrent bien qu'à $m$ égal, le résultat est plus informatif avec le véto qu'avec la pluralité, mais cela mérite une analyse plus poussée. Lorsque la proportion de votes connus est encore assez loin de 1, cette conjecture se vérifie. Ainsi, par exemple, la probabilité qu'il existe au moins un candidat qui ne soit plus gagnant possible (c'est-à-dire que le nombre de gagnants possibles soit $< 10$) devient négligeable à partir de 78 votants pour le véto, mais seulement à partir de 86 votants pour la pluralité. Pour $k = 86$, le nombre de gagnants possibles est presque certainement entre 4 et 7 pour le véto, alors qu'il est encore presque certainement 10 pour la pluralité. Par contre, on observe un retournement de tendance lorsque presque tous les votants ont voté, du moins si l'on se préoccupe de l'aspect binaire (déterminé ou non déterminé) du résultat. Ainsi, la probabilité que le résultat soit déterminé (c'est-à -dire qu'il ne reste qu'un seul gagnant possible) devient non négligeable à partir de 93 votes connus pour la pluralité, mais seulement à partir de 95 ou 96 pour le véto. Pour 98 votes connus, la probabilité que le résultat soit déterminé est aux alentours de 0.4 pour la pluralité contre 0.2 pour le véto. Cependant, même pour ces valeurs de $m$ très élevées, la partie droite de la distribution (probabilité qu'il reste un nombre de gagnants possibles important) reste toujours en faveur du véto, qui présente des distributions moins "aplaties" que la pluralité.

## 3.3 Pluralité à deux tours

La règle de pluralité à deux tours (aussi connu sous le nom de scrutin majoritaire à deux tours) sélectionne dans une première étape les deux candidats ayant les plus grands scores de pluralité (en utilisant comme d'habitude une priorité entre candidats

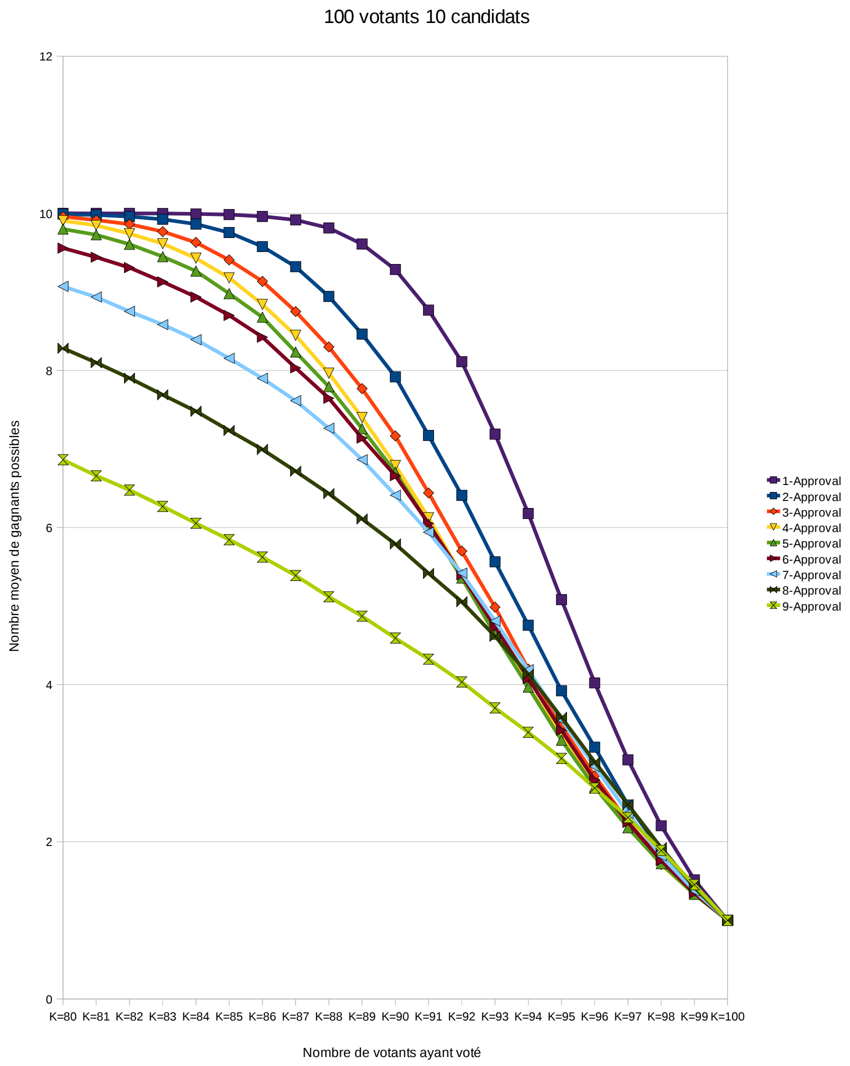


FIG. 1 – Nombre moyen de gagnants possibles (k-approbation)

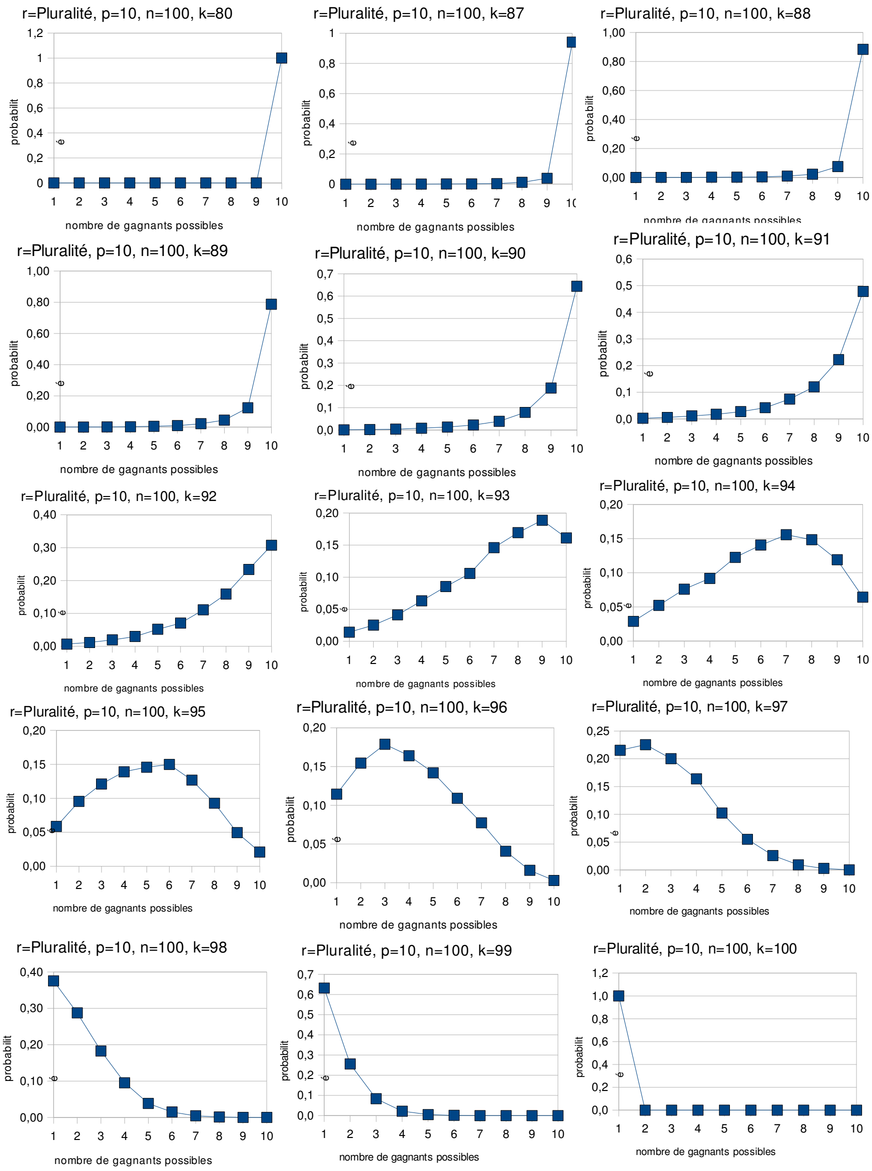


FIG. 2 – Evolution du nombre de gagnants possibles (Pluralité)

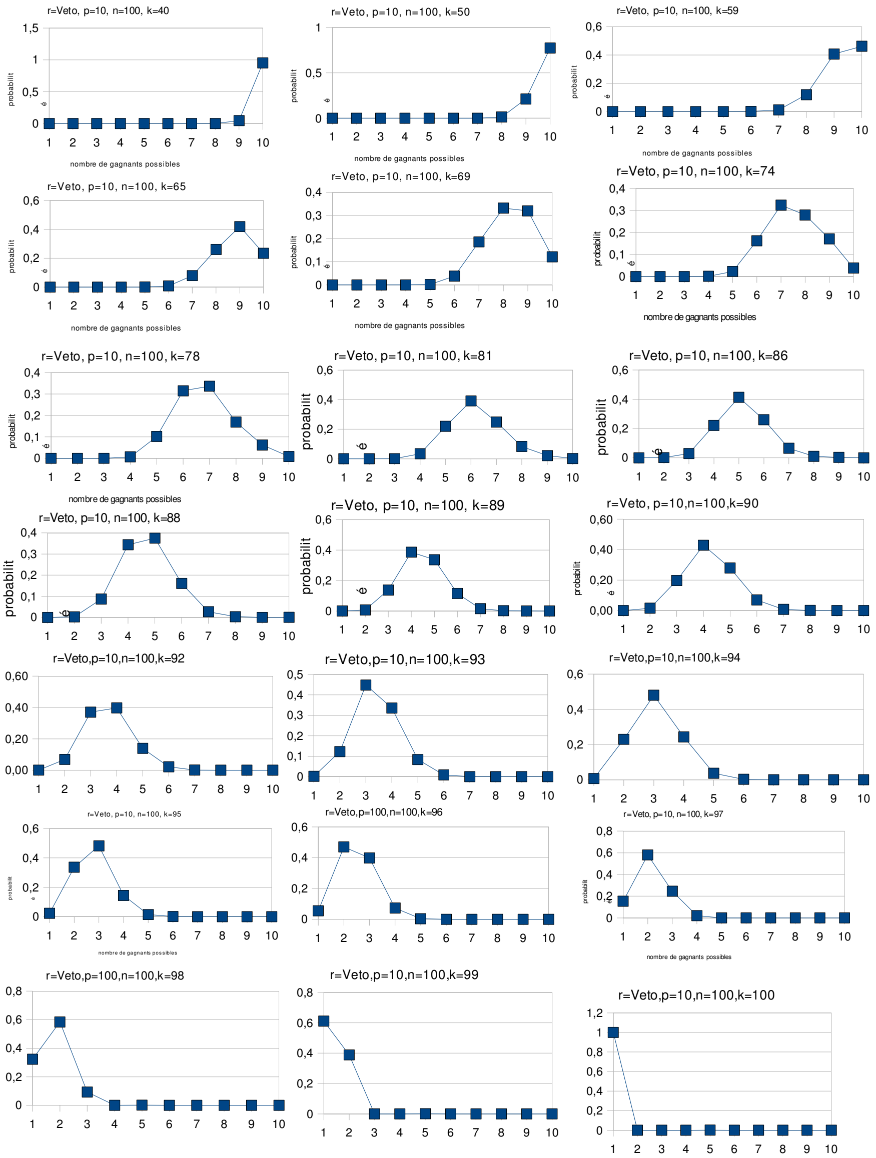


FIG. 3 – Evolution du nombre de gagnants possibles (Veto)

pour départager d'éventuels ex-aequo), et le gagnant est celui des deux qui bat l'autre selon la règle de majorité. L'algorithme 3 de Zuckerman *et al.* (2009) permet le calcul, en temps polynomial, des gagnants possibles pour la pluralité à deux tours.

### 3.4 Borda

La régle de *Borda* est la règle de scoring correspondant au vecteur $\langle p{-}1, p{-}2, \ldots, 0\rangle$.

A ce jour, on ne connaît pas d'algorithme polynomial de calcul des gagnants possibles pour la règle de Borda (on ne sait pas non plus si le problème est NP-difficile – voir Xia *et al.* (2009)). Toutefois, l'algorithme 2 dans Zuckerman *et al.* (2009) nous permet de dire, en temps polynomial, si un candidat est un gagnant possible à une erreur près de 1. Plus précisément, cet algorithme (conçu à l'origine pour déterminer si un profil est manipulable par une coalition de votants), appliqué à un profil $P$, et un candidat $x$, renvoie un entier $q$, avec la garantie que le nombre minimal de nouveaux votants pour lequel $x$ est un gagnant possible (pour Borda) est soit $q$ soit $q-1$.

Dans notre contexte, on peut donc exécuter deux fois l'algorithme pour un profil donné : une fois avec $n-k$ votants, l'algorithme nous donne un sous-ensemble $Poss^-$ ($\subseteq Poss$) des gagnants possibles, et une fois avec un votant de plus, ce qui nous donne cette fois un sur-ensemble $Poss^+$ ($\supseteq Poss$). Lorsque $Poss^+ \setminus Poss^-$ n'est pas vide, il convient pour être précis de vérifier individuellement si ces candidats sont des gagnants possibles ou pas. Pour cela nous allons utiliser un arbre de recherche de la façon suivante : chaque niveau de l'arbre donne le vote d'un nouveau votant, et la profondeur de cet arbre est donc $n-k$, tandis que son facteur de branchement est élevé puisque c'est potentiellement $p!$. On utilisera donc une technique de *backtracking search*, une variante de recherche en profondeur d'abord se contentant d'abord d'étendre un seul fils. Nous allons maintenant tenter de donner des conditions permettant de réduire le facteur de branchement et de ne pas développer une branche non prometteuse. Si on souhaite savoir si le candidat $c_1$ est un gagnant possible alors on peut réduire le facteur de branchement en plaçant $c_1$ en première place pour les $n-k$ votants, et en fixant aux places suivantes tous les candidats ayant un score inférieur ou égal à $c_1$. A chaque niveau de l'arbre, on favorise les votes rapprochant le plus possible le candidat $c_1$ de la tête de l'élection. On peut remarquer que le premier parcours en profondeur correspond exactement à l'algorithme présenté dans Zuckerman *et al.* (2009). Les branches vérifiant la condition suivante ne seront pas développées :

$$\sum_{i \in C \subset X} s(c_i) - s(c_1) > \sum_{i=0}^{|C|-1} (n-k) * (m-1-i)$$

où $s(c_1)$ est le score de Borda du candidat $c_1$, $C$ est l'ensemble des candidats ayant un score de Borda strictement supérieur au score de $c_1$. Cette condition est une condition suffisante pour que le candidat $c_1$ ne soit pas un gagnant possible.

Les résultats des expérimentations sont donnés en Figure 4. On constate ainsi, qu'en moyenne, le nombre de gagnants possibles commence seulement à baisser après que 80% des votants se sont exprimés. Auparavant, *tous* les candidats restent des gagnants possibles (alors que théoriquement, il est possible qu'il n'y ait plus qu'*un* gagnant nécessaire à partir de 2/3 des votes). On remarque aussi, que c'est seulement lorsqu'il ne

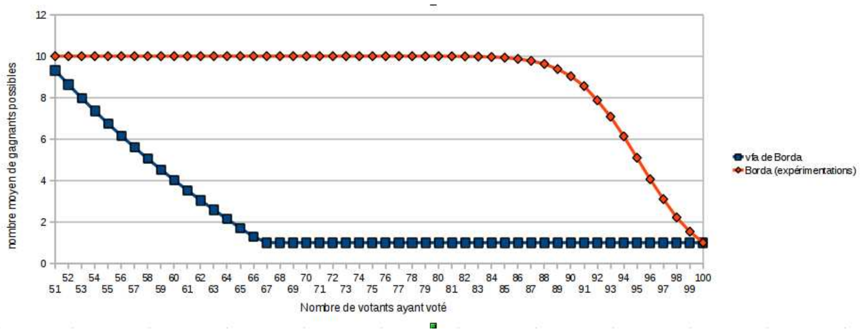


FIG. 4 – Nombre de gagnants possibles (Borda)

reste plus qu'un seul votant que la probabilité d'avoir un seul gagnant possible dépasse les 50%.

## 3.5 Règle de Copeland

Soit $N_P(x, y)$ le nombre de votants dans le profil $P$ qui préfèrent $x$ à $y$.

La règle de *Copeland* élit le candidat qui maximise le score de Copeland défini par $C(x) = |\{y \in X,\ N_P(x, y) > N_P(y, x)\}|$. La règle de Copeland est Condorcet-cohérente.

Pour cette règle pour laquelle le calcul du gagnant possible est $\mathsf{NP}$-difficile, il est intéressant de noter que l'on peut poser le problème comme un *programme linéaire en nombre entiers*. Dénotons par $S_{ij}$ la matrice encodant le graphe de majorité pondéré après que $k$ votants ont exprimé leur vote, et par $\mathcal{V}$ l'ensemble des votants n'ayant pas exprimé leur vote. Posons les $(n-k)(p-1)p$ variables binaires suivantes :

$$x_{ij}^{v} = \begin{cases} 1 & \text{si le votant } v \text{ vote } i \succ j \\ 0 & \text{sinon} \end{cases}$$

Introduisons à présent une nouvelle variable binaire indiquant si le candidat $i$ est préféré au candidat $j$ par une majorité de candidats (c'est-à-dire s'il marque un point au score de Copeland).

$$s_{ij} = \begin{cases} 1 & \text{une majorité stricte de votants préfèrent } i \text{ à } j \\ 0 & \text{sinon} \end{cases}$$

On peut remarquer que l'on considère ici Copeland$_0$[3] puisque dans le cas d'une égalité entre deux candidats $s_{ij}$ sera égal à 0 pour les deux candidats (et non pas 1/2 comme usuellement).

[3] Il existe différentes versions de Copeland selon le nombre de point attribué en cas d'égalité entre deux candidats. Copeland$_\alpha$ $(0 \leq \alpha \leq 1)$ désigne ainsi la règle de Copeland où $\alpha$ points est attribué au candidat $i$ et $j$ si $S_{ij} = S_{ji}$

Chaque votant produisant un ordre linéaire strict et complet, il est clair que chaque votant classe $i$ avant $j$ ou l'inverse, et que la transitivité est respectée, ce qui donne deux contraintes évidentes : (1) et (2). Les contraintes (3) et (4) permettent de lier les variables $x_{ij}^v$ et $s_{ij}$ : si un candidat a plus de $n/2$ point contre un autre candidat, alors nécessairement la valeur de $s_{ij}$ sera à 1. Enfin, nous cherchons un candidat qui soit vainqueur au sens de Copeland, c'est-à-dire qui ait un meilleur score que n'importe quel autre candidat (contrainte (5)). Le programme mathématique résultant est le suivant (où $i$, $j$ et $l$ sont distincts) :

$$\begin{aligned}
\forall i,j \in \mathcal{C}\ \ \forall v \in \mathcal{V} \qquad & x_{ij}^v + x_{ji}^v = 1 && (1)\\
\forall i,j,l \in \mathcal{C}\ \ \forall v \in \mathcal{V} \qquad & x_{ij}^v + x_{jl}^v - x_{il}^v \leq 1 && (2)\\
\forall i,j \in \mathcal{C} \qquad & S_{ij} + \sum_{v=k+1}^{n} x_{ij}^v - \frac{n}{2} \leq \frac{n}{2} s_{ij} && (3)\\
\forall i,j \in \mathcal{C} \qquad & s_{ij} + s_{ji} \leq 1 && (4)\\
\forall i \in \mathcal{C} \qquad & \sum_{j \in \mathcal{C}} s_{cj} - \sum_{j \in \mathcal{C}} s_{ij} \geq 0 && (5)
\end{aligned}$$

Le problème auquel on s'intéresse pour le moment est un problème de *faisabilité* : nous devons vérifier, pour chacun des candidats, s'il vérifie les contraintes (1–5), afin de déterminer s'il est un gagnant possible.

Certaines instances s'avèrent néanmoins difficiles à résoudre par cette méthode, aussi avons-nous développé une approche heuristique alternative, en développant un arbre selon le même principe que celui décrit pour Borda.

L'idée centrale est de borner le score de Copeland des candidats afin de restreindre la recherche. Pour chaque candidat on peut produire un intervalle de valeurs correspondant à leur score de Copeland max. et min. S'il reste $n-k$ votants, il suffit de calculer combien de candidats peuvent encore être rattrapés par $c$, et combien peuvent encore le rattraper.

$$\begin{aligned}
s_{max} &= S_i + \#\{j | N_{ij} + (n-k) > n/2\}\\
s_{min} &= S_i - \#\{j | N_{ji} + (n-k) > n/2\}
\end{aligned}$$

*Exemple.*

Soient 3 candidats $c_1, c_2, c_3$ et 100 votants. 90 votants ont voté et $N(c_1, c_2) = 51$, $N(c_1, c_3) = 46$ et $N(c_2, c_3) = 46$. Sur cette base, on peut donc calculer pour chacun des candidats son score actuel et l'intervalle donné par $s_{min}$ et $s_{max}$. On a donc :

| *candidat* | *score courant* | *score final* |
|---|---|---|
| $c_1$ | 2 | [1,2] |
| $c_2$ | 1 | [0,1] |
| $c_3$ | 0 | [0,2] |

Supposons une règle de départage des ex-æquo suivant l'ordre croissant des candidats. On observe que le candidat $c_1$ est un gagnant possible, et que $c_2$ ne l'est pas (il ne peut pas dépasser le score de $c_1$). Par contre $c_3$ est un gagnant possible ; *et ce alors que son score est plus faible que celui de $c_2$ qui n'est pas possible*. (Pour s'en convaincre, il suffit de considérer que tous les votants votent $c_3 \succ c_1 \succ c_2$.)

On va maintenant formaliser un peu ces notions.

**Définition 1 (Inoffensif)**
*Un candidat $c_1$ est dit* inoffensif *pour un autre candidat $c_2$ si $s_{max}(c_1) < s_{min}(c_2)$ (quoiqu'il arrive, son score de Copeland ne sera jamais plus grand que celui de $c_2$).*

De là, un candidat inoffensif pour au moins un autre candidat n'est pas un gagnant possible. Mais on peut utiliser plus largement cette idée de la manière suivante.

Si $c_1$ est inoffensif pour $c_2$, on peut contraindre la complétion du profil en fixant la place de $c_1$, ce qui diminue le nombre d'ordres à considérer. On peut vérifier que cela consitue une heuristique admissible. En effet, poser cette contrainte n'empêche pas de trouver de solution réalisable. Posons qu'une solution réalisable est trouvée, c'est-à-dire une solution telle que $C(c) > \cdots > C(c_2) > \ldots C(c_1) > \ldots$. Imaginons que cette solution utilise un profil de complétion pour lesquel on a des votes de la forme $v$ :

$$v : c \succ \underbrace{\boxed{c_{t1} \succ \cdots \succ c_{tn}}}_{top} \succ c_2 \succ \underbrace{\boxed{c_{m1} \succ \cdots \succ c_{mn}}}_{midlle} \succ c_1 \succ \underbrace{\boxed{c_{l1} \succ \cdots \succ c_{ln}}}_{last}$$

. Ces votes peuvent être modifiés en $v'$ :

$$v' : c \succ \boxed{c_{t1} \succ \cdots \succ c_{tn}} \succ c_1 \succ c_2 \succ \boxed{c_{m1} \succ \cdots \succ c_{mn}} \succ \boxed{c_{l1} \succ \cdots \succ c_{ln}}$$

On distingue trois blocs de candidats (de taille quelconque) selon leur position relativement à $c_1$ et $c_2$. Pour les candidats des blocs “top” et “last”, les deux votes sont strictement équivalents. Pour les candidats du groupe intermédiaire “middle”, la seule différence est relative à $c_1$ qu'ils battaient, et qui les bat désormais : leur score de Copeland ne peut donc qu'être diminué. Reste le cas de $c_1$ et $c_2$ : le score de $c_1$ peut augmenter mais par définition il reste inférieur à celui de $c_2$, et celui-ci ne peut que baisser. Ainsi si $c$ avait le meilleur score de Copeland, cela restera le cas avec cette transformation : cette solution peut-être trouvée en respectant la contrainte.

Analysons à présent les résultats présentés en Fig.5. La première observation qu'il convient de faire est que cette règle conserve longtemps l'incertitude. Ainsi, avec un seul votant restant, la probabilité d'avoir le résultat établi n'est que de 50%, ce qui est moins que pour les autres règles rencontrées (et environ 1/3 avec deux votes restants). Pour 6 votants restants, on a de grandes chances d'avoir entre 5 et 9 candidats possibles, ce qui reste très élevé.

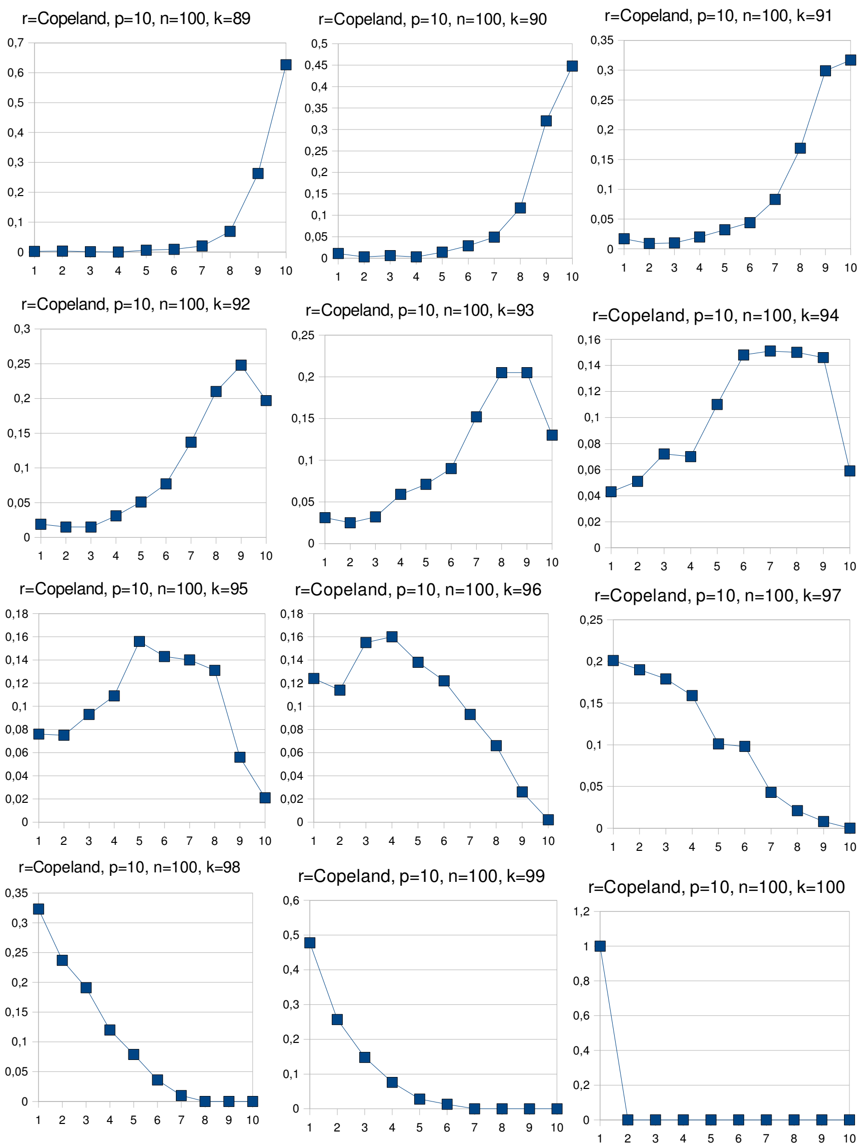


FIG. 5 – Evolution du nombre moyen de gagnants possibles (Copeland)

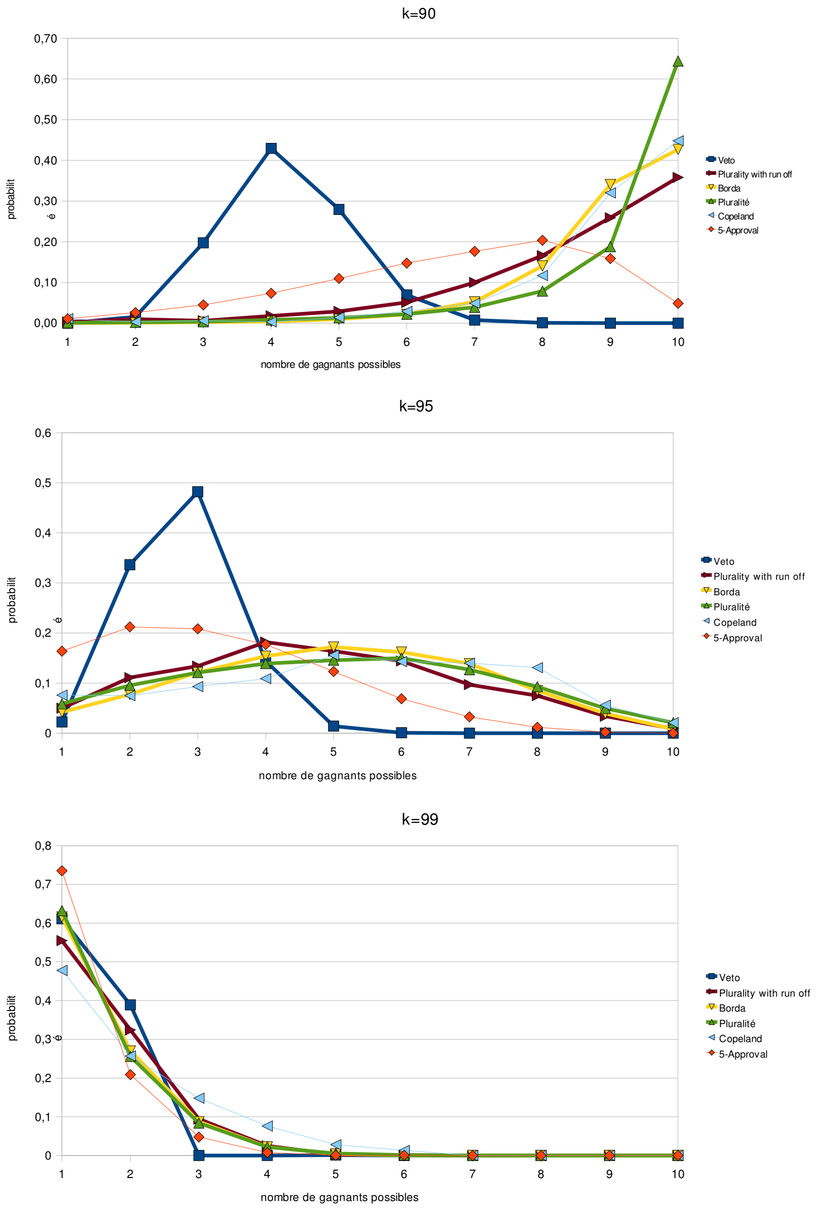


FIG. 6 – Comparaison du nombre moyen de gagnants possibles

## 3.6 Conclusion

Nous avons tenté dans ce papier d'étudier expérimentalement la distribution du nombre de gagnants possibles pour différentes règles de vote. On constate au loyen d'expérimentation que la plupart des règles de votes étudiées gardent le suspense très longtemps. Lorsqu'il manque 20 votants alors que 80 ont voté, rien n'est décidé hormis pour la règle de veto.

Expérimentalement, on constate que la règle de veto se démarque considérablement des autres règles de vote. En effet, si l'on observe sur la figure 6 la distribution du nombre des gagnants possibles pour les règles de vote autres que le veto, on constate que les courbes sont soit concentrées sur $10$ (tous les gagnants sont possibles), soit concentrées sur $1$ (il n'y a qu'un seul gagnant possible), soit ces courbes ont une forme de cloche aplatie, où le nombre de gagnants possibles est indéterminé. La règle de veto, en revanche, a une distribution en forme de cloche pointue. Elle est la seule à pouvoir bloquer efficacement un nombre donné de candidats : par exemple, pour $k = 90$, il n'y a que $3, 4$ ou $5$ gagnants possibles.

Ces observations concordent avec les valeurs de $B(r, p, n)$ de la section 2.3. Pour rappel, cette valeur indique la taille minimale d'une coalition capable d'imposer son candidat (donc, de restreindre le nombre de gagnants possibles à $1$). De *toutes* les règles de scoring, c'est la règle de veto qui possède la valeur de $B$ la plus petite ($B(veto, p, n) = \frac{p-1}{p}$). Parmi les règles de scoring que nous avons envisagées, nous avons $B(veto, p, n) > B(5approbation, p, n) > B(borda, p, n) > B(plurality)$.

On peut aussi remarquer que Copeland est la règle étudiée qui garde l'incertitude le plus longtemps possible.

Pour finir, remarquons que nos expérimentations s'appuient sur l'hypothèse de "culture neutre". Autrement dit, les ordres de chacun des votants sont tirés uniformément et aléatoirement. Lorsque le nombre de votants est important comparé au nombre de candidats, ceci a pour effet d'uniformiser considérablement le profil obtenu. (Pour s'en convaincre, notons que le nombre d'occurences de l'ensemble des ordres lineaires suit une loi multinomiale). Autrement dit, aucun candidat ne sera favorisé par rapport aux autres, et le nombre de gagnants possibles sera d'autant plus important. Il parait donc important pour la suite de nos travaux de comparer ces résultats à ceux basés sur la "culture neutre et anonyme".

# Références


BERG S. & LEPELLEY D. (1992). Note sur le calcul de la probabilité des paradoxes de vote. *Mathématiques et Sciences Humaines*, **120**, 33–48.

CHEVALEYRE Y., LANG J., MAUDET N. & RAVILLY-ABADIE G. (2009). Compiling the votes of a subelectorate. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009)*, p. 97–102.

CONITZER V. & SANDHOLM T. (2002a). Complexity of manipulating elections with few candidates. In *AAAI/IAAI*, p. 314–319.

CONITZER V. & SANDHOLM T. (2002b). Vote elicitation : complexity and strategy-proofness. In *Eighteenth national conference on Artificial intelligence*, p. 392–397, Menlo Park, CA, USA : American Association for Artificial Intelligence.

FRIEDGUT E., KALAI G. & NISAN N. (2008). *Elections Can be Manipulated Often*. Rapport interne, UCLA Department of Economics.

GIBBARD A. (1973). Manipulation of voting schemes : A general result. *Econometrica*, **41**, 587–601.

KONCZAK K. & LANG J. (2005). Voting procedures with incomplete preferences. In *Proc. IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*.

MEIR R., PROCACCIA A. D., ROSENSCHEIN J. S. & ZOHAR A. (2008). Complexity of strategic behavior in multi-winner elections. *J. Artif. Intell. Res. (JAIR)*, **33**, 149–178.

MOULIN H. (1988). *Axioms of Cooperative Decision Making*. Cambridge University Press.

PRITCHARD G. & SLINKO A. (2006). On the average minimum size of a manipulating coalition. *Social Choice and Welfare*, **27**(2), 263–277.

SATTERTHWAITE M. A. (1975). Strategy-proofness and arrow's conditions : Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, **10**(2), 187–217.

SLINKO A. (2006). How the size of a coalition affects its chances to influence an election. *Social Choice and Welfare*, **26**(1), 143–153.

XIA L. & CONITZER V. (2008a). Determining possible and necessary winners under common voting rules given partial orders. In *AAAI*, p. 196–201.

XIA L. & CONITZER V. (2008b). Generalized scoring rules and the frequency of coalitional manipulability. In *EC '08 : Proceedings of the 9th ACM conference on Electronic commerce*, p. 109–118, New York, NY, USA : ACM.

XIA L., ZUCKERMAN M., PROCACCIA A. D., CONITZER V. & ROSENSCHEIN J. S. (2009). Complexity of unweighted coalitional manipulation under some common voting rules. In *The Twenty-First International Joint Conference on Artificial Intelligence (IJCAI 2009)*, p. 348–353, Pasadena, California.

ZUCKERMAN M., PROCACCIA A. D. & ROSENSCHEIN J. S. (2009). Algorithms for the coalitional manipulation problem. *Artif. Intell.*, **173**(2), 392–412.