Approximability and Computational Feasibility of Dodgson's Rule

Supervisors: Dr A. Slinko, Dr G. Pritchard

John C. M^{c}Cabe-Dansted

June 7, 2006

Abstract

Condorcet[1] proposed that a winner of an election is not legitimate unless a majority of the population prefer that alternative to all other alternatives. However such a winner does not always exist. A number of voting rules have been proposed which select the Condorcet winner if it exists, and otherwise selects an alternative that is in some sense closest to being a Condorcet Winner; a prime example is the rule proposed by Dodgson[2](1876).

Unfortunately, Bartholdi et al. (1989) proved that finding the Dodgson winner is an NP-hard problem. Hemaspaandra et al. (1997) refined this result by proving that it is Θ_2^p-complete and hence is not NP-complete unless Θ_2^p=NP. For this reason, we investigate the asymptotic behaviour of approximations to the Dodgson rule as the number of agents gets large.

Under the assumption that all votes are independent and equiprobable, the probability that the Tideman (1987) approximation picks the Dodgson winner does asymptotically converge to 1, but not exponentially fast. We propose a new approximation that does exhibit exponential convergence, and we can quickly verify that it has chosen the Dodgson winner; this allows us to choose the true Dodgson winner with $\mathcal{O}(\ln n)$ expected running time for a fixed number of alternatives m and n agents.

M^c^Garvey (1953) proved that all tournaments are the majority relations for some society. We have proved a generalisation of this theorem for weighted tournaments. We find that this generalisation is useful for simplifying proofs relating to rules which use the weighted majority relation.

Bartholdi et al. (1989) found that we can calculate the Dodgson Score using an ILP that requires no more than $m!m$ variables, we present an improved ILP that requires less than $(m-1)!e$ variables ($e \approx 2.71$). We discover that we can solve this ILP in $\mathcal{O}(\ln n)$ arithmetic operations of $\mathcal{O}(\ln n)$ bits in size. Relaxing the integer constraints results in a new polynomial time rule. In 43 million simulations this new rule failed to pick the

[1]Marie-Jean-Antoine-Nicolas de Caritat, Marquis de Condorcet. (1743–1794)
[2]Charles Lutwidge Dodgson (1832–1898), better known as Lewis Carroll.

Dodgson winner only once, and only given many (25) alternatives. Unlike the Dodgson rule, this rule can break ties in favor of alternatives that are in some sense fractionally better.

We show that Dodgson Score admits no constant error approximation unless P=NP, and admits no Polynomial Time Approximation Scheme (PTAS) for Dodgson Score unless W[2]=FPT.

Acknowledgements

I would first of all like to thank my sister, Kim Dansted, for her suggestions for improving the readability of my introduction and for her general support and assistance while I was going through the grueling process of turning a draft into a legible Thesis.

Dr Arkadii Slinko, my supervisor for much assistance with proof reading. Also, unlike most other Lecturers, he helped guide me through my developmental years, due to his work with the pre-tertiary International Mathematical Olympiad training camps. Indeed, this was why I sought him out as a supervisor.

Dr Geoff Pritchard, my co-supervisor, for assistance with the statistics. Thanks to Pritchard, I have a more rigorous proof that Tideman's approximation converges to Dodgson's rule, which makes proper use of the Multivariate Central Limit Theorem.

Contents

List of Tables

Notations

- x^+: 0 if $x < 0$, x otherwise.
- sgn(x): The sign of x, 1 if $x > 0$, -1 if $x < 0$, 0 if $x = 0$.
- $\lfloor x \rfloor$: Floor of x, the largest integer that is less than or equal to x.
- $\lceil x \rceil$: Ceiling of x, the smallest integer that is greater than or equal to x.
- $|x|$: Absolute value of x, that is x if $x > 0$, -x otherwise.
- $\sum_i$: Summation over each i.
- $\prod_i$: Product over each i.
- $m!$: The factorial of m, i.e. $m! = (1)(2)(3)\cdots(m) = \prod_{i=1}^{m} i$.
- $A \wedge B$: A and B, e.g. A and B will occur.
- $A \vee B$: A or B, e.g. A or B will occur.
- $\neg A$: The negation of A, i.e. the statement "A is false".
- $P(E)$: The probability of event E, a real number in $[0, 1]$.
- $P(A|B)$: The probability of event A will occur given B has occurred or will occur.
- var(X): variance of random variable X.
- $E[X]$: mean of random variable X.
- $A \xrightarrow{D} B$: A converges in distribution to B.
- $N(\mu, \Omega)$: A multivariate normal distribution with means of μ and matrix of correlations Ω.

- $\mathcal{L}(\mathcal{A})$: The set of all linear orders on the set of alternatives A.
- $\mathbb{Z}$: The set of integers.
- $2\mathbb{Z}$: The set of even numbers.
- $\mathbb{N}$: The set $\{1, 2, 3, \ldots\}$ of natural numbers. (also known as positive integers)
- $\mathbb{R}$: The set of real numbers, e.g. $0, 1, \pi, \sqrt{2}$.
- $\mathbb{R}^+$: The set of positive real numbers.
- $\mathbb{R}^k$: The set of real valued k-dimensional vectors.
- $[x, y]$: The range of real numbers from x to y inclusive.
- $[x, y)$: The range of real numbers from x to y, including x but not y.
- $a\mathbf{v}b$: Alternative a is ranked before b in linear order $\mathbf{v}$.
- $\bar{\mathbf{v}}$: The opposite of a linear order $\mathbf{v}$, i.e. $a\mathbf{v}b \Leftrightarrow b\bar{\mathbf{v}}a$.
- n_{ba}: The number of linear orders in our profile $\mathcal{P}$ where alternative b is ranked above alternative a.
- adv(b, a): The advantage of b over a, defined as $(n_{ba} - n_{ab})^+$.
- $\mathbf{1}_k$: A k-dimensional vector with all subscripts equaling 1.
- M^T: The transpose of M, i.e. $M_{ij} \equiv \left(M^T\right)_{ji}$.
- $\#(S)$: The cardinality of the set S, i.e. the number of elements in S.
- $F(x) \in \mathcal{O}(G(x))$: Function F is of order G, i.e. there exists $N \in \mathbb{N}$ and $c > 0$ such that for all n greater than N, $F(x) \leq cG(x)$.
- $\mathbf{x} \leq \mathbf{y}$: vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is less than vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. That is, $x_i \leq y_i$ for all i in $\{1, 2, \ldots, n\}$.
- $\mathbf{x} \geq 0$: vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is positive. That is, $x_i \geq 0$ for all i in $\{1, 2, \ldots, n\}$.
- $A \subseteq B$: A is a subset/submultiset of B.
- x^n: n instances of x, e.g. $\{a, b^2, c\}$ is a multi set with one instance of a and c and two instances of b.
- $\ln x$: The natural log of x, i.e. $\log_e x$.