Approximability of Dodgson's Rule

John C. McCabe-Dansted, Geoffrey Pritchard, Arkadii Slinko

Abstract

It is known that Dodgson's rule is computationally very demanding. Tideman [15] suggested an approximation to it but did not investigate how often his approximation selects the Dodgson winner. We show that under the Impartial Culture assumption the probability that the Tideman winner is the Dodgson winner tends to 1. However we show that the convergence of this probability to 1 is slow. We suggest another approximation — we call it Dodgson Quick — for which this convergence is exponentially fast.

1 Introduction

Condorcet proposed that a winner of an election is not legitimate unless a majority of the population prefer that alternative to all other alternatives. 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 of such as rule was the one proposed by Dodgson [7].

Bartholdi et al. [2] proved that finding the Dodgson winner is, unfortunately, an NP-hard problem. Hemaspaandra et al. [8] refined this result by proving that it is Θ_2^p-complete and hence is not NP-complete unless the polynomial hierarchy collapses. As Dodgson's rule is hard to compute, a number of numerical studies have used approximations [14, 10]. The worst case time required to compute the Dodgson winner from a voting situation is sublinear for a fixed number of alternatives [10], however this algorithm is non-trivial to implement and its running time may grow quickly with the number of alternatives.

We investigate the asymptotic behaviour of simple approximations to the Dodgson rule as the number of agents gets large. Tideman [15] suggested an approximation but did not investigate its convergence to Dodgson. We prove that under the assumption that all votes are independent and each type of vote is equally likely, the probability that the Tideman [15] approximation picks the Dodgson winner asymptotically converges to 1, but not exponentially fast. Although the Simpson rule frequently picks the Dodgson winner [11], it does not converge to Dodgson's rule [10] and is not included in this paper.

We propose a new social choice rule, which we call Dodgson Quick. The Dodgson Quick approximation does exhibit exponential convergence to Dodgson. We may quickly verify that a particular profile has the property that forces the DQ-winner to be the Dodgson winner.

Despite its simplicity, our approximation picked the correct winner in all of 1,000,000 elections with 85 agents and 5 alternatives [10], each generated randomly according to the Impartial Culture assumption. Our approximation can also be used to develop an algorithm to determine the Dodgson winner with $\mathcal{O}(\ln n)$ expected running time for a fixed number of alternatives and n agents.

A result independently obtained by Homan and Hemaspaandra [9] has a lot in common with our result formulated in the previous paragraph, but there are important distinctions as well. They developed a "greedy" algorithm that, given a profile, finds the Dodgson winner with certain probability. Under the Impartial Culture assumption this probability also approaches 1 as we increase the number of agents. However the Dodgson Quick rule is simpler and, unlike their algorithm, the Dodgson Quick rule requires only the information in the weighted majority relation. This makes the Dodgson Quick rule easier to study and compare with other simple rules such as the Tideman rule.

2 Preliminaries

Let A and $\mathcal{N}$ be two finite sets of cardinality m and n respectively. The elements of A will be called alternatives, the elements of $\mathcal{N}$ agents. We assume that the agents have preferences over the set of alternatives represented by (strict) linear orders. By $\mathcal{L}(A)$ we denote the set of all linear orders on A. The elements of the Cartesian product

$$\mathcal{L}(A)^n = \mathcal{L}(A) \times \cdots \times \mathcal{L}(A) \quad (n \text{ times})$$

are called **profiles**. Let $\mathcal{P} = (P_1, P_2, \ldots, P_n)$ be a profile. The linear order P_i represents the preferences of the i^{th} agent; by aP_ib, we denote that this agent prefers a to b. We define n_{xy} to be the number of linear orders in $\mathcal{P}$ that rank x above y, i.e. $n_{xy} = \#\{i \mid xP_iy\}$. A function $W^{\mathcal{P}} \colon A \times A \to \mathbb{Z}$ given by $W^{\mathcal{P}}(a,b) = n_{ab} - n_{ba}$ for all $a, b \in A$, will be called the **weighted majority relation** on $\mathcal{P}$. It is obviously skew symmetric, i.e. $W^{\mathcal{P}}(a,b) = -W^{\mathcal{P}}(b,a)$ for all $a, b \in A$.

Many of the rules to determine the winner use the numbers

$$\operatorname{adv}(a,b) = \max(0, n_{ab} - n_{ba}) = (n_{ab} - n_{ba})^+,$$

which will be called **advantages**. Note that $\operatorname{adv}(a,b) = \max(0, W(a,b)) = W(a,b)^+$ where W is the weighted majority relation on $\mathcal{P}$.

A **Condorcet winner** is an alternative a for which $\operatorname{adv}(b,a) = 0$ for all other alternatives b.

The **Dodgson score** [7, 4, 15], which we denote as $Sc_{\mathbf{d}}(a)$, of an alternative a is the minimum number of neighbouring alternatives that must be swapped to make a a Condorcet winner. We call the alternative(s) with the lowest Dodgson score the **Dodgson winner**(s).

The **Tideman score** [15] $Sc_{\mathbf{t}}(a)$ of an alternative a is

$$Sc_{\mathbf{t}}(a) = \sum_{b \neq a} \operatorname{adv}(b,a).$$

We call the alternative(s) with the lowest Tideman score the **Tideman winner**(s). Tideman [15] suggested the rule based on this score as an approximation to Dodgson.

The **Dodgson Quick (DQ) score** $Sc_{\mathbf{q}}(a)$ of an alternative a, which we introduce in this paper, is

$$Sc_{\mathbf{q}}(a) \quad = \quad \sum_{b\neq a} F(b,a), \text{ where } F(b,a) = \left\lceil \frac{\mathrm{adv}(b,a)}{2} \right\rceil .$$

We call the alternative(s) with the lowest DQ-score the **Dodgson Quick winner**(s) or **DQ-winner**.

The **Impartial Culture** assumption (IC) stipulates that all possible profiles $\mathcal{P} \in \mathcal{L}(A)^n$ are equally likely to represent the collection of preferences of an n-element society of agents $\mathcal{N}$. This assumption does not accurately reflect the voting behaviour of most voting societies and the choice of probability model can affect the similarities between approximations to the Dodgson rule [11]. However the IC is the most simplifying assumption available. As noted by Berg [3], many voting theorists have chosen to focus their research upon the IC. Thus an in depth study of the approximability of Dodgson's rule under the Impartial Culture is a natural first step.

The IC leads to the following $m!$-dimensional multinomial distribution. Let us enumerate all $m!$ linear orders in some way. Let $\mathcal{P} \in \mathcal{L}(A)^n$ be a random profile. Let then X be a vector where each X_i, for $i = 1, 2, \ldots, m!$, represents the number of occurrences of the i^{th} linear order in the profile $\mathcal{P}$. Then, under the IC, the vector X is $(n, k, \mathbf{p})$-multinomially distributed with $k = m!$ and $\mathbf{p} = \mathbf{1}_k/k = (\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$.

Definition 2.1 *A* ***weighted tournament*** *on a set A is any function $W\colon A \times A \to \mathbb{Z}$ satisfying $W(a,b) = -W(b,a)$ for all $a, b \in A$.*

We call $W(a,b)$ the **weight** of an ordered pair of distinct elements (a,b). One can view weighted tournaments as complete directed graphs whose edges are assigned integers characterising the intensity and the sign of the relation between the two vertices that this particular edge connects. The only condition is that if an edge from a to b is assigned integer z, then the edge from b to a is assigned the integer $-z$.

Weighted majority relation $W^{\mathcal{P}}$ on a profile $\mathcal{P}$ defined earlier in this paper is a prime example of a weighted tournament. We say that a profile $\mathcal{P}$ **generates** a weighted tournament W if $W = W^{\mathcal{P}}$. We note that $\mathrm{adv}(a,b) = W^{\mathcal{P}}(a,b)^+$, where $x^+ = \max(0,x)$. Similarly $W^{\mathcal{P}}(a,b) = \mathrm{adv}(a,b) - \mathrm{adv}(b,a)$.

The following theorem generalises the famous M$^{\text{c}}$Garvey theorem [12].

Theorem 2.2 *Let W be a weighted tournament. Then there exists a profile that generates a weighted tournament W if and only if all weights in W have the same parity [5, 13].*

3 Dodgson Quick, A New Approximation

In this section we work under the Impartial Culture assumption.

Definition 3.1 *Let $\mathcal{P} = (P_1, P_2, \ldots, P_n)$ be a profile. We say that the i^{th} agent ranks b* ***directly above*** *a if and only if aP_ib and there does not exist c different from a, b such that aP_ic and cP_ib. We define $D(b, a)$ as the number of agents who rank b directly above a.*

Lemma 3.2 *The probability that $D(x, a) > F(x, a)$ for all x converges exponentially fast to 1 as the number of agents n tends to infinity.*

Proof. As n_{ba} and $D(b, a)$ are binomially distributed with means of $n/2$ and n/m, respectively, from Chomsky's large deviation theorem [6], we know that for a fixed number of alternatives m there exist $\beta_1 > 0$ and $\beta_2 > 0$ s.t.

$$P\left(\frac{D(b,a)}{n} < \frac{1}{2m}\right) \leq e^{-\beta_1 n}, \qquad P\left(\frac{n_{ba}}{n} - \frac{1}{2} > \frac{1}{4m}\right) \leq e^{-\beta_2 n}.$$

We can rearrange the second equation to involve $F(b, a)$,

$$P\left(\frac{n_{ba}}{n} - \frac{1}{2} > \frac{1}{4m}\right) = P\left(\frac{n_{ba} - n_{ab}}{n} > \frac{1}{2m}\right) = P\left(\frac{\mathrm{adv}(b,a)}{n} > \frac{1}{2m}\right).$$

Since $\mathrm{adv}(b, a) \geq F(b, a)$,

$$P\left(\frac{n_{ba}}{n} - \frac{1}{2} > \frac{1}{4m}\right) \quad \geq \quad P\left(\frac{F(b,a)}{n} > \frac{1}{2m}\right).$$

From the law of probability $P(A \vee B) \leq P(A) + P(B)$ it follows that

$$P\left(\frac{F(b,a)}{n} > \frac{1}{2m}\right) \leq e^{-\beta_2 n}, \qquad P\left(\frac{D(b,a)}{n} < \frac{1}{2m}\right) \leq e^{-\beta_1 n},$$

and so for $\beta = \min(\beta_1, \beta_2)$ we obtain

$$P\left(\frac{F(b,a)}{n} > \frac{1}{2m} \text{ or } \frac{D(b,a)}{n} < \frac{1}{2m}\right) \leq e^{-\beta_1 n} + e^{-\beta_2 n} \leq 2e^{-\beta n}.$$

Hence

$$P\left(\exists_x \ \frac{F(x,a)}{n} > \frac{1}{2m} \text{ or } \frac{D(x,a)}{n} < \frac{1}{2m}\right) \quad \leq \quad 2me^{-\beta n}.$$

Using $P(\bar{E}) = 1 - P(E)$, we find that

$$P\left(\forall_x \ \frac{F(x,a)}{n} < \frac{1}{2m} < \frac{D(x,a)}{n}\right) \quad \geq \quad 1 - 2me^{-\beta n}.$$

$\square$

Lemma 3.3 *The DQ-score $Sc_{\boldsymbol{q}}(a)$ is a lower bound for the Dodgson Score $Sc_{\boldsymbol{d}}(a)$ of a.*

Proof. Let $\mathcal{P}$ be a profile and $a \in A$. Suppose we are allowed to change linear orders in $\mathcal{P}$, by repeatedly swapping neighbouring alternatives. Then to make a a Condorcet winner we must reduce $\mathrm{adv}(x, a)$ to 0 for all x and we know that $\mathrm{adv}(x, a) = 0$ if and only if $F(x, a) = 0$. Swapping a over an alternative b ranked directly above a will reduce $n_{ba} - n_{ab}$ by two, but this will not affect $n_{ca} - n_{ac}$ where $a \neq c$. Thus swapping a over b will reduce $F(b, a)$ by one, but will not affect $F(c, a)$ where $b \neq c$. Therefore, making a a Condorcet winner will require at least $\Sigma_b F(b, a)$ swaps. This is the DQ-Score $Sc_{\mathbf{q}}(a)$ of a. □

Lemma 3.4 *If $D(x, a) \geq F(x, a)$ for every alternative x, then the DQ-Score $Sc_{\boldsymbol{q}}(a)$ of a is equal to the Dodgson Score $Sc_{\boldsymbol{d}}(a)$ and the DQ-Winner is equal to the Dodgson Winner.*

Proof. If $D(b, a) \geq F(b, a)$, we can find at least $F(b, a)$ linear orders in the profile where b is ranked directly above a. Thus we can swap a directly over b, $F(b, a)$ times, reducing $F(b, a)$ to 0. Hence we can reduce $F(x, a)$ to 0 for all x, making a a Condorcet winner, using $\Sigma_x F(x, a)$ swaps of neighbouring alternatives. In this case, $Sc_{\mathbf{q}}(a) = \Sigma_b F(b, a)$ is also an upper bound for the Dodgson Score $Sc_{\mathbf{d}}(a)$ of a. Hence $Sc_{\mathbf{q}}(a) = Sc_{\mathbf{d}}(a)$. □

Theorem 3.5 *The probability that the DQ-Score $Sc_{\boldsymbol{q}}(a)$ of an arbitrary alternative a equals the Dodgson Score $Sc_{\boldsymbol{d}}(a)$, converges to 1 exponentially fast.*

Proof. From Lemma 3.4, if $D(x, a) \geq F(x, a)$ for all alternatives x then $Sc_{\mathbf{q}}(a)= Sc_{\mathbf{d}}(a)$. From Lemma 3.2, the probability of this event converges exponentially fast to 1 as $n \to \infty$. □

Corollary 3.6 *The probability that the DQ-Winner is the Dodgson Winner converges to 1 exponentially fast as we increase the number of agents.*

Corollary 3.7 *Suppose that the number of alternatives m is fixed. Then there exists an algorithm that computes the Dodgson score of an alternative a taking as input the frequency of each linear order in the profile $\mathcal{P}$ with expected running time logarithmic with respect to the number of agents (i.e. is $\mathcal{O}(\ln n)$).*

Proof. The are at most $m!$ distinct linear orders in the profile. Hence for a fixed number of alternatives the number of distinct linear orders is bounded. Hence we may find the DQ-score and check whether $D(x, a) \geq F(x, a)$ for all alternatives x using a fixed number of additions. Additions can be performed in time linear with respect to the number of bits and logarithmic with respect to the magnitude of the operands. So we have used an amount of time that is at worst logarithmic with respect to the number of agents.

If $D(x, a) \geq F(x, a)$ for all alternatives x, we know that the DQ-score is the Dodgson score and we do not need to go further. From Lemma 3.2 we know that the probability that we need go further declines exponentially fast, and, if this happens, we can still find the Dodgson score in time polynomial with respect to the number of agents [2]. $\square$

4 Tideman's Rule

In this section we focus our attention on the Tideman rule which was defined in Section 2. We continue to assume the IC.

Lemma 4.1 *Given an even number of agents, the Tideman winner and the DQ-winner will be the same.*

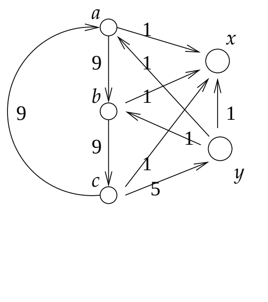
Proof. Since n is even, all weights in the majority relation W are even. Since $\mathrm{adv}(a, b) \equiv W(a, b)^+$ it is clear that all advantages will also be even. Since $\mathrm{adv}(a, b)$ will always be even, $\lceil \mathrm{adv}(a, b)/2 \rceil$ will be exactly half $\mathrm{adv}(a, b)$ and so the DQ-score will be exactly half the Tideman score. Hence the DQ-winner and the Tideman winner will be the same. $\square$

Corollary 4.2 *Let $\mathcal{P}$ be a profile for which the Tideman winner is not the DQ-winner. Then all non-zero advantages are odd.*

Proof. As we must have an odd number of agents, all weights in the majority relation $W^{\mathcal{P}}$ must be odd. Since $\mathrm{adv}(a, b) = W^{\mathcal{P}}(a, b)^+$ the advantage $\mathrm{adv}(a, b)$ must be zero or equal to the weight $W^{\mathcal{P}}(a, b)$. $\square$

Note 4.3 *There are no profiles with three alternatives where the set of DQ-winners and Tideman winners differ. There are profiles with four alternatives where the set of tied winners differ, but no such profile has a unique DQ-winner that differs from the unique Tideman winner [10].*

Example 4.4 *There exist profiles with five alternatives where there is a unique Tideman winner that differs from the unique DQ-winner. By Theorem 2.2, we know we may construct a profile whose weighted majority relation has the following advantages:*



Scores	*a*	*b*	*c*	*x*	*y*
Tideman	*10*	*10*	*9*	*4*	*5*
DQ	*6*	*6*	*5*	*4*	*3*

Here x is the sole Tideman winner, but y is the sole DQ-winner.

Theorem 4.5 *For any $m \geq 5$ there exists a profile with m alternatives and an odd number of agents, where the Tideman winner is not the DQ-winner.*

Example 4.4 demonstrates the existence of a profile with $m = 5$ alternatives for which the Tideman winner is not the Dodgson Quick winner. To extend this example for larger numbers of alternatives, we may add additional alternatives who lose to all of a, b, c, x, y. From Theorem 2.2 there exists a profile with an odd number of agents that generates that weighted majority relation.

Theorem 4.6 *If the number of agents is even, the probability that all of the advantages are 0 does not converge to 0 faster than $\mathcal{O}(n^{-\frac{m!}{4}})$.*

Proof. Let $\mathcal{P}$ be a random profile, $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m!}\}$ be an ordered set containing all $m!$ possible linear orders on m alternatives, and X be a random vector, with elements X_i representing the number of occurrences of $\mathbf{v}_i$ in $\mathcal{P}$. Under the Impartial Culture assumption, X is distributed according to a multinomial distribution with n trials and $m!$ possible outcomes. Let us group the $m!$ outcomes into $m!/2$ pairs $S_i = \{\mathbf{v}_i, \bar{\mathbf{v}}_i\}$. Denote the number of occurrences of $\mathbf{v}$ as $n(\mathbf{v})$. Let the random variable Y_i^1 be $n(\mathbf{v}_i)$ and Y_i^2 be $n(\bar{\mathbf{v}}_i)$. Let $Y_i = Y_i^1 + Y_i^2$.

It is easy to show that, given $Y_i = y_i$ for all i, each Y_i^1 is independently binomially distributed with $p = 1/2$ and y_i trials. It is also easy to show that for an arbitrary integer $n > 0$, a $(2n, 0.5)$-binomial random variable X has a probability of at least $\frac{1}{\sqrt{2n}}$ of equaling n; thus if y_i is even then the probability that $Y_i^1 = Y_i^2$ is at least $\frac{1}{2\sqrt{y_i}}$. Combining these results we get

$$P(\forall_i Y_i^1 = Y_i^2 \mid \forall_i Y_i = y_i \in 2\mathbb{Z}) \quad \geq \quad \prod_i \frac{1}{2\sqrt{y_i}} \geq \prod_i \frac{1}{2\sqrt{n}} = 2^{-\frac{m!}{2}} n^{-\frac{m!}{4}}.$$

It is easy to show that for any k-dimensional multinomially distributed random vector, the probability that all k elements are even is at least 2^{-k+1}; hence the probability that all X_i are even is at least 2^{-k+1} where $k = m!/2$. Hence

$$P(\forall_i X_{i,1} = X_{i,2}) \quad \geq \quad \left(2^{-\frac{m!}{2}+1}\right)\left(2^{-\frac{m!}{2}} n^{-\frac{m!}{4}}\right) = 2^{1-m!} n^{-\frac{m!}{4}}.$$

If for all i, $X_{i,1} = X_{i,2}$ then for all i, $n(\mathbf{v}_i) = n(\bar{\mathbf{v}}_i)$, i.e. the number of each type of vote is the same as its complement. Thus

$$n_{ba} \quad = \quad \sum_{\mathbf{v} \in \{\mathbf{v}: b\mathbf{v}a\}} n(v) = \sum_{\bar{\mathbf{v}} \in \{\bar{\mathbf{v}}: a\bar{\mathbf{v}}b\}} n(\bar{v}) = \sum_{\mathbf{v} \in \{\mathbf{v}: a\mathbf{v}b\}} n(v) = n_{ab},$$

so $\operatorname{adv}(b, a) = 0$ for all alternatives b and a. □

Lemma 4.7 *The probability that the Tideman winner is not the DQ-winner does not converge to 0 faster than $\mathcal{O}(n^{-\frac{m!}{4}})$ as the number of agents n tends to infinity.*

Let $\mathcal{P}$ be a random profile from $\mathcal{L}(A)^n$ for some odd number n. Let $|C|$ be the size of the profile from Theorem 4.5. Let us place the first $|C|$ agents from profile $\mathcal{P}$ into sub-profile C and the remainder of the agents into sub-profile D. There is a small but constant probability that C forms the example from Theorem 4.5, resulting in the Tideman winner of C differing from its DQ-winner. As n, $|C|$ are odd, $|D|$ is even. Thus from Theorem 4.6 the probability that the advantages in D are zero does not converge to 0 faster than $\mathcal{O}(n^{-\frac{m!}{4}})$. If all the advantages in D are zero then adding D to C will not affect the Tideman or DQ-winners. Hence the probability that the Tideman winner is not the DQ-winner does not converge to 0 faster than $\mathcal{O}(n^{-\frac{m!}{4}})$.

Theorem 4.8 *The probability that the Tideman winner is not the Dodgson winner does not converge to 0 faster than $\mathcal{O}(n^{-\frac{m!}{4}})$ as the number of agents n tends to infinity.*

Proof. From Corollary 3.6 the DQ-winner converges to the Dodgson winner exponentially fast. However, the Tideman winner does not converge faster than $\mathcal{O}(n^{-\frac{m!}{4}})$ to the DQ-winner, and hence also does not converge faster than $\mathcal{O}(n^{-\frac{m!}{4}})$ to the Dodgson winner. □

Our next goal is to prove that under the IC the probability that the Tideman winner and Dodgson winner coincide converges asymptotically to 1.

Definition 4.9 *We define the adjacency matrix M, of a linear order $\mathbf{v}$, as follows:*

$$M_{ij} = \begin{cases} 1 & \textit{if} \quad i\mathbf{v}j \\ -1 & \textit{if} \quad j\mathbf{v}i \\ 0 & \textit{if} \quad i = j \end{cases} .$$

Lemma 4.10 *Suppose that $\mathbf{v}$ is a random linear order chosen from the uniform distribution on $\mathcal{L}(A)$. Then its adjacency matrix M is an m^2-dimensional random variable satisfying $E[M] = 0$ and for all $i, j, r, s \in A$:*

$$cov(M_{ij}, M_{rs}) = \begin{cases} 1 & \textit{if} \quad i = r \neq j = s, \\ 1/3 & \textit{if} \quad i = r,\ \textit{but}\, i, j, s \textit{ distinct} \vee j = s,\ \textit{others distinct}, \\ -1/3 & \textit{if} \quad i = s,\ \textit{others distinct} \vee j = r,\ \textit{others distinct}, \\ 0 & \textit{if} \quad i, j, r, s \textit{ distinct} \vee i = j = r = s, \\ -1 & \textit{if} \quad i = s \neq j = r. \end{cases}$$

Proof. Clearly, $E[M_{ij}] = \frac{(1)+(-1)}{2} = 0$. It is well known [1] that $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$ so it follows that $\text{cov}(M_{ij}, M_{rs})E[M_{ij}M_{rs}]$.

Note that for all $i \neq j$ we know that $M_{ii}M_{ii} = 0$, $M_{ij}M_{ij} = 1$, and $M_{ij}M_{ji} = -1$. If $i = r$ and i, j, s are all distinct then the sign of $M_{ij}M_{is}$ for each permutation of i, j and s is as shown below.

$$
\begin{array}{rcccccc}
 & i & i & j & j & s & s \\
 & j & s & i & s & i & j \\
 & s & j & s & i & j & i \\
M_{ij} & + & + & - & - & + & - \\
M_{is} & + & + & + & - & - & - \\
M_{ij}M_{is} & + & + & - & + & - & +
\end{array}
$$

Thus, $E[M_{ij}M_{rs}] = \frac{+1+1-1+1-1+1}{6} = \frac{1}{3}$.

If i, j, r, s are all distinct then there are six linear orders $\mathbf{v}$ where $i\mathbf{v}j$ and $r\mathbf{v}s$, six linear orders $\mathbf{v}$ where $i\mathbf{v}j$ and $s\mathbf{v}r$, six linear orders $\mathbf{v}$ where $j\mathbf{v}i$ and $r\mathbf{v}s$, and six linear orders $\mathbf{v}$ where $j\mathbf{v}i$ and $s\mathbf{v}r$. Hence,

$$E[M_{ij}M_{rs}] \quad = \frac{6(1)(1)+6(1)(-1)+6(-1)(1)+6(-1)(-1)}{24} = 0 \quad .$$

We may prove the other cases for $\text{cov}(M_{ij}, M_{rs})$ in much the same way. $\square$

We note that as $\text{var}(X) = \text{cov}(X, X)$ we also have, $\text{var}(M_{ij}) = 1$ if $i \neq j$, and $\text{var}(M_{ij}) = 0$ if $i = j$.

Define Y to be a collection of random normal variables indexed by i, j for $1 \leq i < j \leq m$ each with mean of 0, and covariance matrix Ω, where

$$\Omega_{ij,rs} \quad = \quad \text{cov}(Y_{ij}, Y_{rs}) = \text{cov}(M_{ij}, M_{rs}),$$

We may use the fact that $i < j$, $r < s$ implies $i \neq j$, $r \neq s$, $(s = i \Rightarrow r \neq j)$ and $(r = j \Rightarrow s \neq i)$ to simplify the definition of Ω as shown below:

$$
\Omega_{ij,rs} \quad = \quad \begin{cases}
1 & \text{if} \quad (r, s) = (i, j), \\
1/3 & \text{if} \quad r = i, s \neq j \text{ or } s = j, r \neq i, \\
-1/3 & \text{if} \quad s = i \text{ or } r = j, \\
0 & \text{if} \quad i, j, r, s \text{ are all distinct.}
\end{cases}
$$

Lemma 4.11 *Let $\mathcal{P} = (P_1, P_2, \ldots, P_n)$ be a profile chosen from the uniform distribution on $\mathcal{L}(A)^n$. Let M_i be the adjacency matrix of P_i. Then, as n approaches infinity, $\sum_{i=1}^{n} M_i/\sqrt{n}$ converges in distribution to*

$$
\begin{bmatrix}
0 & Y_{12} & Y_{13} & \cdots & Y_{1m} \\
-Y_{12} & 0 & Y_{23} & \cdots & Y_{2m} \\
-Y_{13} & -Y_{23} & 0 & \cdots & Y_{3m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-Y_{1m} & -Y_{2m} & -Y_{3m} & \cdots & 0
\end{bmatrix},
$$

where Y is a collection of random normal variables indexed by i, j for $1 \leq i < j \leq m$ each with mean of 0, and covariance matrix Ω, where

$$\Omega_{ij,rs} \quad = \quad cov(Y_{ij}, Y_{rs}) = cov(M_{ij}, M_{rs}).$$

Proof. As $M_1, M_2, \ldots, M_n$ are independent identically-distributed (i.i.d.) random variables, we know from the multivariate central limit theorem [1, p81] that $\sum_{i=1}^{n} M_i/\sqrt{n}$ converges in distribution to the multivariate normal distribution with the same mean and covariance as the random matrix M from Lemma 4.10. As $M^T = -M$ and $M_{ii} = 0$, we have the result. $\square$

Lemma 4.12 *Ω is non-singular.*

Proof. Consider Ω^2 with elements

$$(\Omega^2)_{ij,kl} = \sum_{1\le r<s\le m} \Gamma_{ij,kl}(r,s),$$

where $\Gamma_{ij,kl}(r,s) = \Omega_{ij,rs}\Omega_{rs,kl}$.

If i,j,r,s distinct, then $\Gamma_{ij,ij}(i,j) = 1$ and $\Gamma_{ij,ij}(r,s) = 0$. For $(r,j),(i,s),(r,i),(j,s)$ the function $\Gamma_{ij,ij}$ evaluates to $1/9$.

Let us consider the case $(i,j) = (k,l)$. If $(i,j) = (k,l)$ then

$$\Gamma_{ij,ij}(r,s) = \Omega_{ij,rs}\Omega_{rs,ij} = \begin{cases} (1)^2 & \text{if} \quad (r,s) = (i,j), \\ (1/3)^2 & \text{if} \quad r = i, s \neq j \text{ or } s = j, r \neq i, \\ (-1/3)^2 & \text{if} \quad s = i, (r \neq j) \text{ or } r = j, (s \neq i), \\ 0 & \text{if} \quad i,j,r,s \text{ are all distinct.} \end{cases}$$

Recall that $r < s$, $i < j$ and $r,s \in [1,m]$. Let us consider for how many values of (r,s) each of the above cases occur:

- $(r,s) = (i,j)$: This occurs for exactly one value of (r,s).
- $r = i, s \neq j$: Combining the fact that $r < s$ and $r = i$ we get $i < s$. Thus $s \in (i,j) \cup (j,m]$, and there are $(j-i-1)+(m-j) = (m-i-1)$ possible values of s. As there is only one possible value of r this means that there are also $(m-i-1)$ possible values of (r,s).
- $s = j, r \neq i$: Combining the fact that $r < s$ and $s = j$ we get $r < j$. Thus $r \in [1,i) \cup (i,j)$, and there are $(i-1)+(j-i-1) = (j-2)$ possible values of (r,s).
- $s = i$: Here we want $r \neq j$, however $r < s = i < j$, so explicitly stating $r \neq j$ is redundant. Combining the fact that $r < s$ and $s = i$ we get $r < i$. Hence $r \in [1,i]$ and there are $i-1$ possible values for (r,s).
- $r = j$: Here we want $s \neq i$, however $i < j = r < s$, so explicitly stating that $r \neq j$ is redundant. From here on we will not state redundant inequalities. Combining the fact that $r < s$ and $r = j$ we get $j < s$. Hence $s \in (j,m]$ and there are $m-j$ possible values for (r,s).

Hence,

$$\begin{aligned}\sum_{1\le r<s\le m}\Gamma_{ij,ij}(r,s) &= 1+(m+j-i-3)\left(\frac{1}{9}\right)+(m+i-j-1)\left(\frac{1}{9}\right)\\ &= \left(9+(m+j-i-3)+(m+i-j-1)\right)/9=\frac{2m+5}{9}.\end{aligned}$$

Let us consider now the case $i=k,\ j\neq l$. Then

$$\Gamma_{ij,il}(r,s) = \begin{cases} (1)(1/3) = 1/3 & \text{if } (i,j)=(r,s),\\ (1/3)(1) = 1/3 & \text{if } r=i, s=l\neq j,\\ (1/3)(1/3) = 1/9 & \text{if } r=i, s\neq j, s=\neq l,\\ (1/3)(0) = 0 & \text{if } s=j\neq l, r\neq i,\\ (-1/3)(-1/3) = 1/9 & \text{if } s=i,\\ (-1/3)(1/3) = -1/9 & \text{if } r=j, s=l,\\ (-1/3)(0) = 0 & \text{if } r=j, s\neq l,\\ 0 = 0 & \text{if } i,j,r,s \text{ are all distinct,}\end{cases}$$

hence,

$$\begin{aligned}\sum_{1\le r<s\le m}\Gamma_{ij,il}(r,s) &= \frac{1}{3}+\frac{1}{3}+\sum_{1\le r<s\le m, r=i, s\neq j, s=\neq l}\frac{1}{9}+\sum_{1\le r<s\le m, s=i}\frac{1}{9}-\frac{1}{9}\\ &= \frac{1}{3}+\frac{1}{3}+\sum_{i<s\le m}\frac{1}{9}-\frac{2}{9}+\sum_{1\le r<i}\frac{1}{9}-\frac{1}{9}\\ &= \frac{1}{3}+(m-i)\frac{1}{9}+(i-1)\frac{1}{9}=\frac{m+2}{9}.\end{aligned}$$

Similarly for $i\neq k,\ j=l$, we may show $(\Omega^2)_{ij,kj}=\frac{m+2}{9}$. If $j=k$ then

$$(\Omega^2)_{ij,kl} = -\frac{1}{3}-\frac{1}{3}+\frac{1}{9}-\sum_{1\le r<i, r\neq i}\frac{1}{9}-\sum_{j<s\le m, s\neq l}\frac{1}{9}=-\frac{m+2}{9},$$

similarly for $l=i$. If i,j,k,l are all distinct, $(\Omega^2)_{ij,kl}$ equals 0. Consequently

$$\Omega^2 = \left(\frac{m+2}{3}\right)\Omega-\left(\frac{m+1}{9}\right)I.$$

Since the matrix Ω satisfies $\Omega^2=\alpha\Omega+\beta I$ with $\beta\neq 0$ it has an inverse, hence Ω is not singular. □

Theorem 4.13 *The probability that the Tideman winner and Dodgson winner coincide converges asymptotically to* 1 *as* $n\to\infty$.

Proof. We will prove that the Tideman winner asymptotically coincides with the Dodgson Quick winner. The Tideman winner is the alternative $a\in A$ with the minimal value of

$$G(a) = \sum_{b\in A}\operatorname{adv}(b,a),$$

while the DQ-winner has minimal value of

$$F(a) \quad = \quad \sum_{b \in A} \left\lceil \frac{\mathrm{adv}(b,a)}{2} \right\rceil .$$

Let a_T be the Tideman winner and a_Q be the DQ-winner. Note that $G(c)-m \leq 2F(c) \leq G(c)$ for every alternative c. If for some b we have $G(b) - m > G(a_T)$, then $2F(b) \geq G(b)-m > G(a_T) \geq 2F(a_T)$ and so b is not a DQ-winner. Hence, if $G(b) - m > G(a_T)$ for all alternatives b distinct from a_T, then a_T is also the DQ-winner a_Q. Thus,

$$P(a_T \neq a_Q) \leq P(\exists_{a \neq b} |G(a) - G(b)| \leq m) = P\left(\exists_{a \neq b} \left| \frac{G(a) - G(b)}{\sqrt{n}} \right| \leq \frac{m}{\sqrt{n}}\right).$$

It follows that for any $\epsilon > 0$ and sufficiently large n, we have

$$P(a_T \neq a_Q) \quad \leq \quad P\left(\exists_{a \neq b} \left| \frac{G(a) - G(b)}{\sqrt{n}} \right| \leq \epsilon\right).$$

We will show that the right-hand side of the inequality above converges to 0 as n tends to ∞. All probabilities are non-negative so $0 \leq P(a_T \neq a_Q)$. From these facts and the sandwich theorem it will follow that $\lim_{n\to\infty} P(a_T \neq a_Q) = 0$.

Let

$$G_j \quad = \quad \sum_{i<j} (Y_{ij})^+ + \sum_{k>j} (-Y_{jk})^+ ,$$

where variables Y_{ij} come from the matrix (1) to which $\sum_{i=1}^{n} M_i/\sqrt{n}$ converges by Lemma 4.11. Thus,

$$\lim_{n\to\infty} P\left(\exists_{a \neq b} \left| \frac{G(a) - G(b)}{\sqrt{n}} \right| \leq \epsilon\right) \quad = \quad P\left(\exists_{i \neq j} |G_i - G_j| \leq \epsilon\right)$$

Since $\epsilon > 0$ is arbitrary,

$$\lim_{n\to\infty} P(a_T \neq a_Q) \quad \leq \quad P(\exists_{i \neq j} G_i = G_j).$$

For fixed $i < j$ we have

$$G_i - G_j = -Y_{ij} + \sum_{k<i} (-Y_{ki})^+ + \sum_{k>i, k\neq i} (Y_{ik})^+ - \sum_{k<j, k\neq i} (Y_{kj})^+ - \sum_{k>j} (-Y_{jk})^+ .$$

Define v so that $G_i - G_j = -Y_{ij} + v$. Then $P(G_i = G_j) = P(Y_{ij} = v) = E[P(Y_{ij} = v \,|\, v)]$. Since Y has a multivariate normal distribution with a non-singular covariance matrix Ω, it follows that $P(Y_{ij} = v \,|\, v) = 0$. That is, $P(G_i = G_j) = 0$ for any i, j where $i \neq j$. Hence $P(\exists_{i \neq j} G_i = G_j) = 0$. As discussed previously in this proof, we may now use the sandwich theorem to prove that $\lim_{n\to\infty} P(a_T \neq a_Q) = 0$. $\square$

5 Conclusion

In this paper we showed that, under the Impartial Culture assumption, the Tideman rule converges to the Dodgson's rule when the number of agents tends to infinity. However we discovered that a new rule, which we call Dodgson Quick, approximates Dodgson's rule much better and converges to it much faster. The Dodgson Quick rule is computationally very simple, however in our simulations [10] it picked the Dodgson winner in all of 1,000,000 elections with 85 agents and 5 alternatives.

These results, the simplicity of Dodgson Quick's definition and the ease with which its winner can be computed make Dodgson Quick a highly effective tool for theoretical and numerical study of Dodgson's rule under the Impartial Culture assumption. Despite the popularity of the Impartial Culture as a simplifying assumption, it is highly unrealistic and our theorems do not apply if the slightest deviation from impartiality occurs. Our previous numerical results [11] suggest that introduction of homogeneity into the random sample may cause these approximations to diverge from the Dodgson rule. The most interesting question for further research, that this paper rises, is whether or not the Dodgson Quick rule approximates Dodgson's rule under the Impartial Anonymous Culture assumption and other models for the population.

While there is no significant difference in the difficulty of computing the Dodgson Quick winner or the Tideman winner, the Tideman rule can be easier to reason with in some circumstances. We find that the Tideman rule is often useful to study properties of the Dodgson rule where rapid convergence is not required.

References

[1] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. John Wileys and Sons, Brisbane, 2nd edition, 1984.

[2] Bartholdi, III., Tovey, C. A., and Trick, M. A. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare: Springer-Verlag*, 6:157–165, 1989.

[3] Berg, S. Paradox of voting under an urn model: The effect of homogeneity. *Public Choice*, 47:377–387, 1985.

[4] Black, D. *Theory of committees and elections*. Cambridge University Press, Cambridge, 1958.

[5] Debord, B. *Axiomatisation de procédures d'agrégation de préférences*. Ph.D. thesis, Université Scientifique, Technologique et Médicale de Grenoble, 1987. `http://tel.archives-ouvertes.fr/tel-00010237/en/`.

[6] Dembo, A. and Zeitouni, O. *Large deviations techniques*. Johns and Barlett, 1993.

[7] Dodgson, C. L. *A method for taking votes on more than two issues*. Clarendon Press, Oxford, 1876. Reprinted in (Black, 1958) with discussion.

[8] Hemaspaandra, E., Hemaspaandra, L., and Rothe, J. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, 44(6):806–825, 1997.

[9] Homan, C. M. and Hemaspaandra, L. A. Guarantees for the success frequency of an algorithm for finding dodgson-election winners. Technical Report Technical Report TR-881, Department of Computer Science, University of Rochester, Rochester, NY, 2005. `https://urresearch.rochester.edu/retrieve/4794/tr881.pdf`.

[10] M^c^Cabe-Dansted, J. C. *Feasibility and Approximability of Dodgson's rule*. Master's thesis, Auckland University, 2006. `http://dansted.org/thesis06`.

[11] M^c^Cabe-Dansted, J. C. and Slinko, A. Exploratory analysis of similarities between social choice rules. *Group Decision and Negotiation*, 15:1–31, 2006. `http://dx.doi.org/10.1007/s00355-005-0052-4`.

[12] M^c^Garvey, D. C. A theorem on the construction of voting paradoxes. *Econometrica*, 21:608–610, 1953.

[13] Ratliff, T. C. A comparison of dodgson's method and kemeny's rule. *Social Choice and Welfare*, 18(1):79–89, 2001. ISSN 0176-1714 (Print) 1432-217X (Online). `http://dx.doi.org/10.1007/s003550000060`.

[14] Shah, R. *Statistical Mappings of Social Choice Rules*. Master's thesis, Stanford University, 2003.

[15] Tideman, T. N. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4:185–206, 1987.

John C. M^c^Cabe-Dansted
School of Computer Science and Software Engineering
M002 The University of Western Australia
35 Stirling Highway, Crawley 6009, Western Australia
Email: `gmatht@gmail.com`

Dr Geoffrey Pritchard
Department of Statistics, Auckland University
Private Bag 92019, Auckland 1142, New Zealand
Email: `geoff@stat.auckland.ac.nz`

Dr Arkadii Slinko
Department of Mathematics, Auckland University
Private Bag 92019, Auckland 1142, New Zealand
Email: `a.slinko@auckland.ac.nz`