Soc Choice Welfare (2008) 31:311–330
DOI 10.1007/s00355-007-0282-8

ORIGINAL PAPER

Approximability of Dodgson's rule

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

Received: 12 September 2006 / Accepted: 17 October 2007 / Published online: 1 December 2007

Abstract It is known that Dodgson's rule is computationally very demanding. Tideman (Soc Choice Welf 4:185–206, 1987) 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 converges to 1 as the number of voters increase. However we show that this convergence is not exponentially fast. We suggest another approximation—we call it Dodgson Quick—for which this convergence is exponentially fast. Also we show that the Simpson and Dodgson rules are asymptotically different.

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. 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 of such as rule is the rule that was proposed by Dodgson (1876).

J. C. McCabe-Dansted (✉)
School of Computer Science and Software Engineering, University of Western Australia,
35 Stirling Highway, Crawley, WA 6009, Australia
e-mail: john@csse.uwa.edu.au

G. Pritchard
Department of Statistics, Auckland University, Auckland, New Zealand
e-mail: geoff@stat.auckland.ac.nz

A. Slinko
Department of Mathematics, University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: a.slinko@auckland.ac.nz

Bartholdi et al. (1989) proved that finding the Dodgson winner is, unfortunately, 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 the polynomial hierarchy collapses. As Dodgson's rule is hard to compute, it is important to have simple and fast approximations to it. Shah (2003) would not have included a statistical analysis of Dodgson's rule if Tideman's approximation was not available.

We investigate the asymptotic behaviour of simple approximations to the Dodgson rule as the number of agents gets large. Tideman (1987) 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 Impartial Culture (IC) assumption), the probability that the Tideman (1987) approximation picks the Dodgson winner asymptotically converges to 1, but not exponentially fast.

We propose a new social choice rule, which we call Dodgson Quick. The Dodgson Quick approximation does exhibit exponential convergence to Dodgson. The proof of this is based on the discovery of a large class of profiles for which it is certain that the Dodgson Quick winner is also the Dodgson winner. As it is possible to verify whether a profile is a member of this class in polynomial time, this provides us with an easy way to verify that the Dodgson Quick approximation has picked the same winner. Even if we only consider results that can be verified in this way to be correct, the proof still demonstrates exponential convergence. This, together with its simplicity and other nice properties, makes our new approximation useful in computing 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 (McCabe-Dansted 2006), 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 (2005) 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 to other simple rules such as the Tideman rule. Also unlike their rule, the scores of the Dodgson Quick rule are lower bounds for the Dodgson scores; this makes the Dodgson Quick scores useful for quickly identifying alternatives that cannot be Dodgson winners (McCabe-Dansted, 2006).

Our experimental results (McCabe-Dansted and Slinko 2006) showed that Simpson's and Dodgson's rules are very close. However, in the present paper we discover that under the Impartial Culture assumption, the frequency that the Simpson rule picks the Dodgson winner does not converge to one.

Whereas this paper defines closeness in terms of probabilities, other papers have defined closeness such that two close rules will always pick similar rankings of the alternatives; under this definition of closeness it is hard to find an approximation for the Dodgson rule. For scoring rules we may rank the alternatives according to their scores. Klamler (2004) found that there exist profiles such that the Borda rule picks the reverse ranking to the Dodgson ranking. Ratliff (2001, 2002) found that the Dodgson winner may appear in any position in the ranking of the Kemeny rule and of any positional rule. It is not known whether there exists any polynomial time approximation to the Dodgson rule that always chooses a similar ranking to the Dodgson ranking.

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**. The profiles represent the collection of preferences of an n-element society of agents $\mathcal{N}$. A family of mappings $F = \{F_n\}$, $n \in \mathbb{N}$,

$$F_n \colon \mathcal{L}(A)^n \to A,$$

is called a **social choice function** (SCF).

Let $\mathcal{P} = (P_1, P_2, \ldots, P_n)$ be a profile. If a linear order $P_i \in \mathcal{L}(A)$ represents the preferences of the ith agent, then by $a P_i b$, where $a, b \in A$, we denote that this agent prefers a to b. We say that the ith agent ranks b **directly above** a if and only if $a P_i b$ and there does not exist c different from a, b such that $a P_i c$ and $c P_i b$. We say that a pair of alternatives is neighbouring iff a is ranked directly above b or b is ranked directly above a. We define n_{xy} to be the number of linear orders in $\mathcal{P}$ that rank x above y, i.e., $n_{xy} = \#\{i \mid x P_i y\}$. The approximations we consider depend upon the information contained in the matrix $N_{\mathcal{P}}$, where $(N_{\mathcal{P}})_{ab} = n_{ab}$. 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

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

which will be called **advantages**.

A **Condorcet winner** is an alternative a for which $\mathrm{adv}(b, a) = 0$ for all other alternatives b. A Condorcet winner does not always exist. The rules we consider below attempt to pick an alternative that is in some sense closest to being a Condorcet winner. These rules will always pick the Condorcet winner when it exists; such rules are called Condorcet consistent rules.

The social choice rules we consider are based on calculating the vector of **scores** and the alternative with the lowest score wins. Let the lowest score be s. It is possible that more than one alternative has a score of s. In this case we may have a set of winners with cardinality greater than one. Strictly speaking, to be a social choice function, a rule has to output a single winner. Rules are commonly modified to achieve this by splitting ties. One of the most popular methods of splitting ties is to split ties according to the preferences of the first agent. However we will usually study the set of tied winners rather than the single winner output from a tie-breaking procedure, as this will allow us to empirically detect smaller differences between the rules.

The **Dodgson score** (Dodgson 1876, see e.g., Black 1958, Tideman 1987), which we denote as $S_{\mathbf{d}}(a)$, 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 **Simpson score** (Simpson 1969, see e.g., Laslier 1997) $S_{\mathbf{s}}(a)$ of an alternative a is

$$S_{\mathbf{s}}(a) = \max_{b \neq a} \operatorname{adv}(b, a).$$

We call the alternative(s) with the lowest Simpson score the **Simpson winner**(s). That is, the alternative with the smallest maximum defeat is the Simpson winner. This is why the rule is often known as the Maximin or Minimax rule.

The **Tideman score** (Tideman 1987) $S_{\mathbf{t}}(a)$ of an alternative a is

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

We call the alternative(s) with the lowest Tideman score the **Tideman winner**(s). Tideman (1987) suggested the rule based on this score as an approximation to Dodgson.

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

$$S_{\mathbf{q}}(a) = \sum_{b \neq a} F(b, a),$$

where

$$F(b, a) = \left\lceil \frac{\operatorname{adv}(b, a)}{2} \right\rceil.$$

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

The difference between the DQ and Tideman rules may appear trivial. Simply dividing the advantages by two would not result in different winner. The DQ and Tideman winners only differ because the result is rounded up. Despite the similarity between the DQ and Tideman rules, DQ is a much better approximation. The DQ scores are usually exactly the same as the Dodgson scores (Theorem 5). This makes the DQ rule a very close approximation to the Dodgson rule. Although the Tideman rule is quite close to the Dodgson rule (McCabe-Dansted and Slinko 2006), it may favour an alternative that is defeated by many other alternatives by small margins where the Dodgson rule would not (see Example 2). This small difference is enough to prevent the Tideman rule from converging exponentially quickly (Theorem 17).

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 $\mathcal{N}$, i.e., all agents are independent and they choose their linear orders from the uniform distribution on $\mathcal{L}(A)$. This assumption is of course does not accurately reflect the voting behaviour of most voting societies. Worse, we have found that the choice of probability model for the population can affect the similarities between approximations to the Dodgson rule (McCabe-Dansted and Slinko 2006). However the IC is the most simplifying assumption available. As noted by Berg (1985), 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 assumption is a natural first step.

The Impartial Culture assumption 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 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 $\mathbf{p} = \mathbf{1}_k/k = (\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$.

3 Dodgson Quick, a new approximation

In this section, we work under the Impartial Culture assumption.

Definition 1 We define $D(b, a)$ as the number of agents who rank b directly above a, in our profile $\mathcal{P}$.

Lemma 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 (see Dembo and Zeitouni 1993), we know that for a fixed number of alternatives m there exist $\beta_1 > 0$ and $\beta_2 > 0$ such that

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

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

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

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

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

Thus we have

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

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

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

Hence

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

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

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

Lemma 3 *The DQ-score $S_{\mathbf{q}}(a)$ is a lower bound for the Dodgson Score $S_{\mathbf{d}}(a)$ 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 $S_{\mathbf{q}}(a)$ of a.

Lemma 4 *If $D(x, a) \geq F(x, a)$ for every alternative x, then the DQ-Score $S_{\mathbf{q}}(a)$ of a is equal to the Dodgson Score $S_{\mathbf{d}}(a)$ 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, $S_{\mathbf{q}}(a) = \Sigma_b F(b, a)$ is an upper bound for the Dodgson Score $S_{\mathbf{d}}(a)$ of a. From Theorem 3 above, $S_{\mathbf{q}}(a)$ is also a lower bound for $S_{\mathbf{d}}(a)$. Hence $S_{\mathbf{q}}(a) = S_{\mathbf{d}}(a)$.

Theorem 5 *The probability that the DQ-Score $S_{\mathbf{q}}(a)$ of an arbitrary alternative a equals the Dodgson Score $S_{\mathbf{d}}(a)$, converges to* 1 *exponentially fast.*

Proof From Lemma 4, if $D(x, a) \geq F(x, a)$ for all alternatives x then $S_{\mathbf{q}}(a)= S_{\mathbf{d}}(a)$. From Lemma 2, the probability of this event converges exponentially fast to 1 as $n \to \infty$.

Corollary 6 *The probability that the DQ-Winner is the Dodgson Winner converges to* 1 *exponentially fast as we increase the number of agents.*

Although Bartholdi et al. (1989) showed that for a fixed number of alternatives, the amount of time required to compute the Dodgson winner is polynomial, the order of this polynomial can be large and the order grows with the number of alternatives. Below it will be shown that the Dodgson winner can be found in expected time that is logarithmic with respect to the number of agents, under the IC assumption.

Corollary 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. Bartholdi et al. (1989) proposed an algorithm for finding the Dodgson score in a polynomial amount of time; the expected amount of time spent in this algorithm declines to zero because we know that the probability that $D(x, a) < F(x, a)$ for some alternative x declines exponentially fast (Lemma 2).

Corollary 8 *There exists an algorithm that computes the Dodgson winner taking as input the frequency of each linear order in the profile $\mathcal{P}$ with expected running time that is logarithmic with respect to the number of agents.*

4 Tideman's rule

In this section, we focus our attention on the Tideman rule which was defined in Sect. 2. We continue to use the Impartial Culture assumption.

Lemma 9 *Given an even number of agents, the Tideman winner and the DQ-winner will be the same.*

Proof Since n is even, $n_{ab} - nba$ is even for each pair of alternatives a and b. From this, and $\mathrm{adv}(a, b) \equiv (n_{ab} - n_{ba})^+$ 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.

Corollary 10 *Let $\mathcal{P}$ be a profile for which the Tideman winner is not the DQ-winner. Then the profile has an odd number of agents and all non-zero advantages are odd.*

Lemma 11 *There is no profile with three alternatives such that the Tideman winner is not the DQ-winner.*

Proof The Tideman and Dodgson Quick rules both pick the Condorcet winner when it exists, so if a Condorcet winner exists the Tideman winner and DQ-winner will be the same. It is well known that the absence of a Condorcet winner on three alternatives means that we can rename these alternatives a, b and c so that $\mathrm{adv}(a, b) > 0$, $\mathrm{adv}(b, c) > 0$, and $\mathrm{adv}(c, a) > 0$. These advantages must be odd from the previous corollary. Hence there exist integers $i, j, k \in \mathbb{Z}$ such that $\mathrm{adv}(a, b) = 2i - 1$, $\mathrm{adv}(b, c) = 2j - 1$, and $\mathrm{adv}(c, a) = 2k - 1$. The DQ-Scores and Tideman scores of a, b, c are k, i, j and $2k - 1,\ 2i - 1,\ 2j - 1$ respectively. From here the result is clear, since if $i > j > k$ then $2i - 1 > 2j - 1 > 2k - 1$.

Lemma 12 *For a profile with four alternatives there does not exist a pair (a, b) of alternatives such that a is a DQ-winner but not a Tideman winner, and b is a Tideman winner but not a DQ-winner.*

Proof By way of contradiction assume that such alternatives a, b exist. Consider the relationship between the Tideman score $S_{\mathbf{t}}(c)$ and the DQ-score $S_{\mathbf{q}}(c)$ of some alternative c:

$$\begin{aligned} S_{\mathbf{t}}(c) &= \sum_{d \in A} \mathrm{adv}(d, c) = 2 \sum_{d \in A} \left\lceil \frac{\mathrm{adv}(d, c)}{2} \right\rceil - \#\{d : \mathrm{adv}(d, c) \notin 2\mathbb{Z}\} \\ &= 2S_{\mathbf{q}}(c) - \#\{d : \mathrm{adv}(d, c) \notin 2\mathbb{Z}\}. \end{aligned}$$

Since the set of Tideman winners and DQ-winners differ, n must be odd and hence all non-zero advantages must be odd. It follows that:

$$\#\{d : \mathrm{adv}(d, c) \notin 2\mathbb{Z}\} = \#\{d : \mathrm{adv}(d, c) > 0\}$$

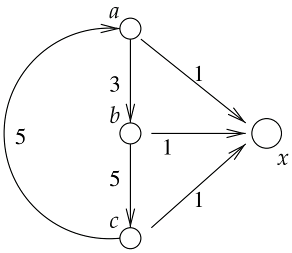
As the Tideman winner is not the DQ winner, there is no Condorcet winner, and so for each alternative c there are one to three alternatives d such that $\mathrm{adv}(d, c) > 0$. Thus, $2S_{\mathbf{q}}(c) - 3 \le S_{\mathbf{t}}(c) \le 2S_{\mathbf{q}}(c) - 1$, and, in particular,

$$S_{\mathbf{t}}(a) \le 2S_{\mathbf{q}}(a) - 1, \quad 2S_{\mathbf{q}}(b) - 3 \le S_{\mathbf{t}}(b).$$

Given that a is DQ-winner and b is not, we know that $S_{\mathbf{q}}(a) \le S_{\mathbf{q}}(b) - 1$. Thus by substitution, $S_{\mathbf{t}}(a) \le 2(S_{\mathbf{q}}(b) - 1) - 1 = 2S_{\mathbf{q}}(b) - 3 \le S_{\mathbf{t}}(b)$. This shows that if b is a Tideman winner, so is a. By contradiction the result must be correct.

Theorem 13 *There exists a profile that generates a weighted tournament W if and only if all weights in W have the same parity* (*Debord* 1987, *see also Klamler* 2004).

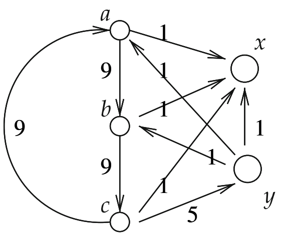
Example 1 There do exist profiles with four alternatives where the set of tied Tideman winners differs from the set of tied DQ-winners. By Theorem 13, we know we may construct a profile whose weighted majority relation has the following advantages:



Scores	a	b	c	x
Tideman	5	3	5	3
DQ	3	2	3	3

Here x, b are tied Tideman winners, but b is the sole DQ-winner.

Example 2 There do exist profiles with five alternatives where there is a unique Tideman winner that differs from the unique DQ-winner. By Theorem 13, 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 14 *For any $m \geq 5$ there exists a profile with m alternatives and an odd number of agents, where the unique Tideman winner is not the unique DQ-winner.*

In Example 2 we gave an example 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 by an odd margin of defeat.

Theorem 15 *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

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

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

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

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

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

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

Lemma 16 *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 14. 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 14, resulting in the Tideman winner of C differing from its DQ-winner. As n and $|C|$ are odd, $|D|$ is even. Thus from Theorem 15 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 17 *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 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.

Below it will be shown that, under the Impartial Culture assumption, the probability that the Tideman winner and Dodgson winner coincide converges asymptotically to 1.

Definition 18 We define the adjacency matrix M, of a linear order $\mathbf{v}$, as follows:

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

Lemma 19 *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 the following equations for all* $i, j, r, s \in A$.

$$\begin{aligned} E[M] &= 0 \\ \text{cov} M_{ij} M_{rs} &= E[M_{ij} M_{rs}] \\ &= \begin{cases} 1 & \text{if } \ i = r \neq j = s, \\ 1/3 & \text{if } \ i = r, \ \textit{but } i, j, s \textit{ distinct} \ \vee \ j = s, \ \textit{others distinct}, \\ -1/3 & \text{if } \ i = s, \ \textit{others distinct} \ \vee \ j = r, \ \textit{others distinct}, \\ 0 & \text{if } \ i, j, r, s \textit{ distinct} \ \vee \ i = j \ \vee \ r = s, \\ -1 & \text{if } \ i = s \neq j = r. \end{cases} \end{aligned}$$

Proof Clearly, $E[M_{ij}] = \frac{(1)+(-1)}{2} = 0$. As $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$ (see e.g., Walpole and Myers 1993, p. 97), $\text{cov}(M_{ij}, M_{rs}) = E[M_{ij} M_{rs}] - (0)(0) = E[M_{ij} M_{rs}]$. 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.

	i	i	j	j	s	s
	j	s	i	s	i	j
	s	j	s	i	j	i
M_{ij}	+	+	−	−	+	−
M_{is}	+	+	+	−	−	−
$M_{ij} M_{is}$	+	+	−	+	−	+

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,

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

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

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

Example 20 For example, for $m = 4$ the covariances with M_{12} are shown in the matrix

$$\mathfrak{L} = \begin{bmatrix} 0 & 1 & 1/3 & 1/3 \\ -1 & 0 & -1/3 & -1/3 \\ -1/3 & 1/3 & 0 & 0 \\ -1/3 & 1/3 & 0 & 0 \end{bmatrix},$$

where $\mathfrak{L}_{ij} = \mathrm{cov}(M_{ij}, M_{12})$.

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

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

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:

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

Lemma 21 *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

$$\Omega_{ij,rs} = \mathrm{cov} Y_{ij} Y_{rs} = \mathrm{cov} M_{ij} M_{rs}.$$

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 (see e.g., Anderson 1984; p. 81) 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 19. As $M^T = -M$ and $M_{ii} = 0$, we have the result.

Lemma 22 *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

$$\begin{aligned}
\Gamma_{ij,ij}(i, j) &= \Omega_{ij,ij}\Omega_{ij,ij} = (1)(1) = 1,\\
\Gamma_{ij,ij}(r, j) &= \Omega_{ij,rj}\Omega_{rj,ij} = (1/3)(1/3) = 1/9,\\
\Gamma_{ij,ij}(i, s) &= \Omega_{ij,is}\Omega_{is,ij} = (1/3)(1/3) = 1/9,\\
\Gamma_{ij,ij}(r, i) &= \Omega_{ij,ri}\Omega_{ri,ij} = (-1/3)(-1/3) = 1/9,\\
\Gamma_{ij,ij}(j, s) &= \Omega_{ij,js}\Omega_{js,ij} = (-1/3)(-1/3) = 1/9,\\
\Gamma_{ij,ij}(r, s) &= \Omega_{ij,rs}\Omega_{ij,rs} = 0.
\end{aligned}$$

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

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

Recall that $r < s$, $i < j$ and $r, s \in [1, m]$. 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).
- $r = i, s \ne j$: 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).
- $s = j, r \ne i$: 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 $r \ne j$, however $r < s = i < j$, so explicitly stating $r \ne j$ 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 $s \ne i$, however $i < j = r < s$, so explicitly stating that $s \ne i$ 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,

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

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

$$\Gamma_{ij,il}(r,s) = \Omega_{ij,rs}\Omega_{rs,il} = \begin{cases} 1\Omega_{rs,il} & \text{if } (r,s) = (i,j), \\ {}^1\!/\!{}_3\Omega_{rs,il} & \text{if } r = i, s \neq j \text{ or } s = j, r \neq i, \\ -{}^1\!/\!{}_3\Omega_{rs,il} & \text{if } s = i \text{ or } r = j, \\ 0 & \text{if } i, j, r, s \text{ are all distinct.} \end{cases}$$

More precisely,

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

hence,

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

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

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

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 23 *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

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

while the DQ-winner has minimal value of

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

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,

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

thus for any $\epsilon > 0$ and sufficiently large n, we have

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

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

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

where variables Y_{ij} come from the matrix (1) to which $\sum_{i=1}^{n} M_i/\sqrt{n}$ converges by Lemma 21. Thus,

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

Since $\epsilon > 0$ is arbitrary,

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

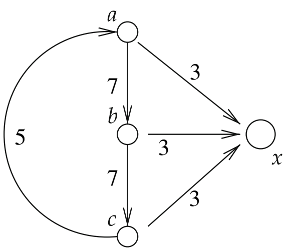
For fixed $i < j$ we have

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

Define v so that $G_i - G_j = -Y_{ij} + v$. Then $P(G_i = G_j) = P(Y_{ij} = v) = E[P(Y_{ij} = v \mid v)]$. Since Y has a multivariate normal distribution with a non-singular covariance matrix Ω, it follows that $P(Y_{ij} = v \mid v) = 0$. 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 $\lim_{n\to\infty} P(a_T \neq a_Q) = 0$.

In Sect. 5 numerical results are presented which show that the limiting probability of the Simpson rule and Tideman winners differ under the IC assumption is not zero. An outline of a formal proof that these rules do not converge is presented below:

Lemma 24 *The probability that the Simpson and Tideman winners differ under IC does not converge to zero as $n \to \infty$ under the IC assumption.*



Scores	a	b	c	x
Simpson	5	7	7	3
Tideman	5	7	7	9

Table 1 Number of occurrences per 10,000 Elections with 5 alternatives that the Dodgson winner was not chosen

Voters	3	5	7	9	15	17	25	85	257	1025
DQ	1.5	1.9	1.35	0.55	0.05	0.1	0	0	0	0
Tideman	1.5	2.3	2.7	3.95	6.05	6.85	7.95	8.2	5.9	2.95
Simpson	57.6	65.7	62.2	57.8	48.3	46.6	41.9	30.2	23.4	21.6

Table 2 Number of occurrences per 10,000 Elections with 5 alternatives that the set of Dodgson winners is not chosen

Voters	3	5	7	9	15	17	25	33	85	257	1025
DQ	4.31	4.41	3.21	1.94	0.27	0.08	0.04	0	0	0	0
Tideman	4.31	5.57	7.31	8.43	12.73	13.15	15.46	16.35	15.18	10.2	5.4

We see that a is the unique Tideman winner, while x is the Simpson winner. From Lemma _ we know that the limiting distribution of $\frac{n_{ab}-n_{ba}}{\sqrt{n}}$ for each unordered pair of alternatives $\{a, b\}$ converges to a non-degenerate multivariate normal distribution as $n \to \infty$ under the IC assumption. There exists a point on this multivariate distribution where advantages have the same ratios as above. Clearly there exists an open neighborhood around that point where the unique Tideman winner differs from the unique Simpson winner. Therefore, the limiting probability that the Tideman winner differs from the Simpson winner is not zero.

5 Numerical results

In this section, we present Tables 1 to 5 demonstrating the rate of convergence to Dodgson of the Dodgson Quick rule introduced in this paper in comparison to the Tideman rule. These tables show that the convergence of the Tideman winner to the Dodgson Winner occurs much slower than the exponential convergence of the DQ-Winner. We also study the asymptotic limit of the probability that the Simpson winner is the Dodgson winner as we increase the number of agents. The C, MATLAB and R code used to generate these tables is availiable upon request from the authors.

In these 10,000 simulations we were breaking ties according to the preferences of the first agent. In Table 2 we present the results of another 10,000 simulations in which we consider the rules as social choice correspondences and do not break ties.

Another question is how well does Dodgson Quick approximate the Dodgson rule when the number of alternatives is different from 5 or when the number of agents is not large in comparison to the number of agents. From Table 3, it appears that the DQ-approximation is still reasonably accurate under these conditions. This table was generated by averaging 10,000 simulations, and splitting ties according to the preferences of the first agent.

To give meaning to these figures, let us compare them with the figures in Tables 4 and 5. We see that even where the number of agents is not very large, the Dodgson Quick rule seems to do a slightly better job of approximating the Dodgson

Table 3 Frequency that the DQ-winner is the Dodgson winner

No. of alternatives	No. of agents						
	3	5	7	9	15	25	85
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0.9984	0.9976	0.9980	0.9992	0.9999	1.0000	1.0000
7	0.9902	0.9875	0.9879	0.9933	0.9980	0.9995	1.0000
9	0.9792	0.9742	0.9778	0.9837	0.9924	0.9978	0.9999
15	0.9468	0.9327	0.9338	0.9412	0.9571	0.9743	0.9988
25	0.8997	0.8718	0.8661	0.8731	0.8971	0.9265	0.9840

Table 4 Frequency that the Tideman winner is the Dodgson winner

No. of alternatives	No. of agents						
	3	5	7	9	15	25	85
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0.9984	0.9974	0.9961	0.9972	0.9936	0.9917	0.9930
7	0.9902	0.9864	0.9852	0.9868	0.9845	0.9805	0.9847
9	0.9792	0.9730	0.9724	0.9731	0.9718	0.9760	0.9815
15	0.9468	0.9292	0.9263	0.9273	0.9379	0.9485	0.9649
25	0.8997	0.8691	0.8620	0.8625	0.8833	0.9113	0.9534

Table 5 Frequency that the Simpson winner is the Dodgson winner

No. of alternatives	No. of agents						
	3	5	7	9	15	25	85
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0.9433	0.9307	0.9339	0.9398	0.9493	0.9575	0.9714
7	0.8734	0.8627	0.8689	0.8786	0.9018	0.9153	0.9404
9	0.8256	0.8153	0.8167	0.8251	0.8562	0.8808	0.9124
25	0.5895	0.5772	0.6147	0.6322	0.7114	0.7529	0.7957

rule than Tideman's approximation. We also see that Simpson's rule does a particularly poor job of approximating the Dodgson winner when the number of alternatives is large.

It appears that Simpson's rule is not a very accurate approximation of Dodgson's Rule. The probability that the Simpson winner does not equal the Dodgson winner is much greater than for Tideman or DQ. We may ask, does the Simpson rule eventually converge to the Dodgson rule as we increase the number of voters, and, if not, how close does it get?

From Lemma 2 and Theorem 23 we know that the Dodgson winner, Dodgson Quick winner, and Tideman winner all asymptotically converge as we increase the number

Table 6 Number of occurrences per 1,000 Elections that the Simpson winner is not the Dodgson winner. (Limit as $n \to \infty$)

No. of alternatives	3	4	5	6	7	8	9	10	11	15	25	50
#(DO $\neq$ SI)	0	7.14	17.47	28.7	39.9	50.4	60.7	70.2	79.2	109	160	225

of agents. Hence we may compute the asymptotic probability that the Simpson winner is equal to the Dodgson winner, by computing the asymptotic probability that the Simpson winner equals the Tideman winner.

From Lemma 21 we know that the matrix of advantages converges to a multivariate normal distribution as we increase the number of agents. We have generated 10,000,000 samples from this distribution for each value of m to produce Table 6.

Note that as the number of agents approaches infinity, the probability of a tie approaches 0, and so tie breaking is irrelevant in this table. In Table 6, we see that even with an infinite number of voters, the Simpson rule is not especially close to the Dodgson rule.

6 Conclusion

In this paper we showed that under the Impartial Culture assumption the Tideman rule approximates Dodgson's rule and converges to it, 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. We also show that Simpson's rule does not converge to Dodgson's rule asymptotically despite often selecting the same winner. The Dodgson Quick rule is computationally very simple. However, in our simulations, Dodgson Quick picked the Dodgson winner in all of 1,000,000 elections with 85 agents and 5 alternatives. We give numerical results illustrating the rate of convergence of Dodgson Quick to Dodgson.

These results, the simplicity of Dodgson Quick's definition and the ease with which its winner can be computed make Dodgson Quick an 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 unrealistic and our theorems do not apply if the slightest deviation from impartiality occurs. Our previous numerical results (McCabe-Dansted and Slinko 2006) suggest that if homogeneity is introduced this 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.

References

Anderson TW (1984) An introduction to multivariate statistical analysis, 2nd edn. Wiley, Brisbane

Bartholdi III, Tovey CA, Trick MA (1989) Voting schemes for which it can be difficult to tell who won the election. Soc Choice Welf 6:157–165

Berg S (1985) Paradox of voting under an urn model: the effect of homogeneity. Public Choice 47: 377–387

Black D (1958) Theory of committees and elections. Cambridge University Press, Cambridge

Debord B (1987) Caractérisation des matrices des préférences nettes et méthodes d'agrégation associées. (Characterization of net preference matrices and related aggregation methods). Math Sci Hum 97: 5–17

Dembo A, Zeitouni O (1993) Large deviations techniques. Johns and Barlett

Dodgson CL (1876) A method for taking votes on more than two issues. Clarendon Press, Oxford. Reprinted in Black (1958) with discussion

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

Homan CM, Hemaspaandra LA (2005) 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. http://urresearch.rochester.edu/retrieve/4794/tr881.pdf

Klamler C (2004) The Dodgson ranking and the Borda count: a binary comparison. Math Soc Sci 48(1):103–108. http://ideas.repec.org/a/eee/matsoc/v48y2004i1p103-108.html

Laslier J-F (1997) Tournament solutions and majority voting. Springer, Berlin

McCabe-Dansted JC (2006) Feasibility and approximability of Dodgson's rule. Master's thesis, Auckland University. http://dansted.org/thesis06

McCabe-Dansted JC, Slinko A (2006) Exploratory analysis of similarities between social choice rules. Group Decis Negot 15:1–31. http://dx.doi.org/10.1007/s00355-005-0052-4

Ratliff TC (2001) A comparison of Dodgson's method and Kemeny's rule. Soc Choice Welf 18(1):79–89. ISSN 0176-1714 (Print) 1432-217X (Online). http://dx.doi.org/10.1007/s003550000060

Ratliff TC (2002) A comparison of Dodgson's method and the Borda count. Econ Theory 20(2):357–372. ISSN 0938-2259 (Print) 1432-0479 (Online). http://dx.doi.org/10.1007/s001990100218

Shah R (2003) Statistical mappings of social choice rules. Master's thesis, Stanford University

Simpson PB (1969) On defining areas of voter choice: Professor tullock on stable voting. Q J Econ 83(3): 478–490

Tideman TN (1987) Independence of clones as a criterion for voting rules. Soc Choice Welf 4:185–206

Walpole RE, Myers RH (1993) Probability and statistics for engineers and scientists. Maxwell Macmillian International, Sydney