Stein's Method: Expository Lectures and Applications
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 46 (2004) 75–92

5. The stationary distribution in the antivoter model: exact sampling and approximations

Mark Huber[1] **and Gesine Reinert**[2]

Duke University and University of Oxford

Abstract: The antivoter model is a Markov chain on regular graphs which has a unique stationary distribution, but is not reversible. This makes the stationary distribution difficult to describe. Despite the fact that in general we know nothing about the stationary distribution other than it exists and is unique, we present a method for sampling exactly from this distribution. The method has running time $O(n^3 r/c)$, where n is the number of nodes in the graph, c is the size of the minimum cut in the graph, and r is the degree of each node in the graph. We also show that the original chain has $O(n^3 r/c)$ mixing time. For the antivoter model on the complete graph we derive a closed form solution for the stationary distribution. Moreover we bound the total variation distance between the stationary distribution for the antivoter model on a multipartite graph and the stationary distribution on the complete graph, using Stein's method. Finally, we present computational experiments comparing the empirical Stein's method for estimating the stationary distribution to the classical frequency estimate.

5.1. The antivoter model

The antivoter model is a variant of the well studied voter model (See Liggett [8]). In the voter model, a graph is given a two coloring. Independently with exponential waiting times, nodes randomly change their color to match a random neighbor. In finite graphs, the voter model will eventually result in all nodes having the same color.

The antivoter model is a variant where again nodes are updated independently with exponential waiting times between updates. However, in the antivoter model the color of the node is changed to the opposite of a randomly chosen neighbor. Matloff [9] first introduced the antivoter model on infinite lattices. The case of finite graphs was examined by Donnelly and Welsh [5], and by Aldous and Fill [2]. Unlike the voter model, none of the states of the antivoter model are absorbing, and in fact most of the states are strongly connected.

We consider a discrete time version of this model, run as follows. At each time step, a node of the graph is selected uniformly at random among all nodes. A neighbor of the chosen node is then selected uniformly at random from the set of neighboring nodes. The color of our selected node is changed to be the opposite of the selected neighbor; we use the colors 0 and 1 (another typical convention is to label the colors by -1 and 1). That this chain is not reversible is easy to see – if a node colored 0 is entirely surrounded by nodes which are also colored 0, picking that node always results in its color changing to 1, with no chance of moving back to 0.

[1]Department of Mathematics and the Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0320, USA. e-mail: mhuber@math.duke.edu

[2]Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK. e-mail: reinert@stats.ox.ac.uk

We restrict our attention to graphs G which are regular of degree r. Assume that G is neither bipartite nor an n-cycle. Donnelly and Welsh [5] were able to show that under these assumptions, the antivoter model is an ergodic chain over the set of colorings of the graph in which both colors are used at least once. Moreover the stationary distribution is unique, and its support is the set of all nonunanimous configurations.

One statistic of particular interest to us is W, which is the number of nodes colored 1. By symmetry $E[W] = n/2$. Prompted by a question posed by Aldous, Rinott and Rotar [12] were able to show that the distribution of W for regular graphs is approximately normal given certain conditions on the graph, as the degree of the graph goes to infinity.

On the complete graph, this model can be viewed as a special case of the Moran model [10] in genetics. Assume a population of fixed size n consists of individuals of two types, 0 and 1, say. As pointed out by Donnelly [4], the Moran model can be formulated as follows. Individuals die at rate one, independently of all other events. When an individual dies, then, with probability α_1 it is replaced by an individual of type 1, with probability α_2 it is replaced by an individual of type 0, and with probability $1 - \alpha_1 - \alpha_2$ it is replaced by the offspring of an individual chosen at random from the members of the population present immediately before the death occurs. In this case the offspring will always be of the same allelic type as its parent. To see the relation with the antivoter model, suppose we choose a vertex v at random. In the antivoter model, a random neighbor is chosen, and the color is changed to the opposite $\bar{y}$ of y, the color of the random neighbor. As on the complete graph all vertices are neighbors, the probability for vertex v to retain color x is proportional to the number of vertices of color $\bar{x}$ in the graph. Thus the antivoter model is a special case of the Moran model, with $\alpha_1 = \alpha_2 = 0$.

In this paper we present several results about the antivoter model. First, we give an efficient algorithm for sampling exactly from the stationary distribution of the chain, and show that the chain mixes rapidly. Yet estimating the stationary distribution from exact samples can be rather time consuming, in particular for the tails of the distribution. In the special case of the antivoter model on the complete graph, we are able to give a closed form solution for the stationary distribution. Intuitively, if the class sizes in the multipartite graph are not too large, the multipartite graph resembles a complete graph, and thus the stationary distributions should be similar. We make this intuition precise; using Stein's method, we derive a bound on the approximation of the stationary distribution of W for multipartite graphs by the stationary distribution for complete graphs. Finally, we implement the empirical Stein's method for this chain, and present evidence that it computes more accurate estimates of the stationary distribution with fewer samples. This is illustrated on multipartite graphs, as we have the normal distribution as the known limiting regime for the distribution of W; however, the method is easily applicable to other graphs as well.

5.2. Exact sampling

Our method for exactly sampling from the stationary distribution of this chain uses the coupling from the past (CFTP) methodology. CFTP rests on having a complete coupling chain. Recall that a chain is a coupling chain if whenever two particles on the chain enter the same state at time t, they remain in that state for all $T \geq t$. That is, if X and Y are two particles on the chain, $X_t = Y_t \Rightarrow X_T = Y_T$ for all $T \geq t$.

Suppose that we start a particle on every state of the chain. If all of these particles have coupled by time t, we say that the chain has completely coupled. CFTP uses a complete coupling chain as a subroutine to obtain an exact sample from the stationary distribution of the chain. The running time is of the same order as the expected time for the chain to completely couple.

Complete coupling also provides a bound on the mixing time of the chain, as measured by the total variation distance. The total variation between two distributions p and π is $\|p-\pi\|_{TV} = \sup_A |p(A)-\pi(A)|$ (where the supremum is taken over all measurable sets A). Let $p^t_{x_0}$ denote the distribution of X_t given that $X_0 = x_0$. Then for $\epsilon > 0$, the mixing time of the chain is the smallest value of t such that $\|p - \pi\|_{TV} \leq \epsilon$. A well known theorem of Doeblin's [3, 1] tells us that the total variation distance between $p^t_{x_0}$ and π is bounded about by the probability that complete coupling has not occurred. Therefore, if we show that complete coupling has occcurred by time t with high probability, then the chain will mix well by that time.

Given that complete coupling gives us an exactly stationary sample, one might inquire as to why we are also interested in bounding the mixing time, which only gives approximate samples. Often, simulators will not be interested in obtaining a single sample, but many samples. One way to get these samples is to take an entire set of values from a sample path $X_0, \ldots, X_t$. These value will be correlated, but knowledge of the mixing time of the chain allows us to bound this correlation. Of course, we should always start such a sample path with X_0 stationary, and so our perfect sampling techniques are also useful here.

In the antivoter model on a graph with n nodes, the state space is of size 2^n. It is impractical to keep track of 2^n different particles, and so we employ an idea to keep track of all states that reduces finding a stationary state in the antivoter chain to detecting absorbtion in the voter chain.

To keep track of the nodes, we keep track of variables $x_{i,t}$, where $x_{i,t}$ is the color of node i at time t. Now suppose that at the tth time step, we choose node i to have color the opposite of node j. Then $x_{i,t+1} = 1 - x_{j,t}$. Suppose that we are running the chain forward from some starting time T It is easy to see via induction that at some time step T_0, each x_{i,T_0} will either be $x_{j,T}$ or $1 - x_{j,T}$ for some j. This is a version of the coalescent coupling; node j could be viewed as the most recent ancestor of all the nodes.

At this point, the state of the entire chain depends only on what the value of $x_{j,T}$ is, 0 or 1. Once that is known, then the entire state of the chain is known. To determine this, we first note another move that preserves the stationary distribution of this Markov chain, namely, at each step we swap the color classes with probability 1/2. That is, we flip a fair coin at each step, and if it comes up heads, each node colored 1 becomes color 0 and vice versa. This update step actually only needs to be done once, since we only need keep track of the parity of the number of times that the coin came up heads, and adjust the state accordingly at the end.

This allows us to do the following. Suppose at time t, all of the nodes have value $x_{j,T}$ or $1 - x_{j,T}$. Then at time $t - 1$, we must have flipped a coin. With probability 1/2, $x_{j,T} = 0$ and with probability 1/2, $x_{j,T} = 1$. Either way, we can now write down the state of the chain at time t.

This is a complete coupling–no matter what the state of the chain was at time $T - 1$, at time t it has moved to the same state. Together with coupling from the past (CFTP), this gives us an algorithm for exact sampling from the chain with running time of the same order as the expected time the chain takes to completely

couple.

So what is the running time of this procedure? The time needed is the time for one variable $x_{j,T}$ to "spread" out to hit all of the other nodes. But this is exactly the time needed for a node in the voter model to influence every other node of the graph. The running time of this procedure will be exactly the same as the absorbtion time of the voter model with n colors. We bound this time below.

Theorem 5.1. *The expected time the above chain takes to completely couple is bounded above by $n^3 r/c$, where r is the degree of the graph and c is the minimum cut.*

(Note that $c \geq r$ always, so that the expression makes sense.) Our proof technique will be to show that the number of known nodes at each time step is a submartingale. Therefore it will be useful to have some facts about submartingales available.

In particular, the Upcrossing Inequality (see Port [11]) will be useful in determining the expected time until complete coupling.

Lemma 5.1 (Upcrossing Inequality). *Suppose $X_1, \ldots$ is a submartingale (so $E[X_{t+1} - X_t | X_t] \geq 0$). Let τ be a stopping time, and $U(\alpha, \beta)$ be the number of times that $X_i \leq \alpha \leq \beta \leq X_{i+1}$. Then*

$$EU(\alpha, \beta) \leq \frac{1}{\beta - \alpha}[E(X_\tau - \alpha)^+ - E(X_1 - \alpha)^+].$$

The following corollary will be needed.

Corollary 5.1. *Suppose $X_0, \ldots$ is a submartingale taking values in $\{0, \ldots, n\}$ with $X_0 = 0$, $\tau = \inf_t X_t = n$, $|X_{t+1} - X_t| \leq 1$, and $P(X_{t+1} \neq X_t) \geq \epsilon_{X_t} > 0$. Then*

$$E[\tau] \leq \sum_{i=0}^{n-1} \frac{2(n-i)+1}{\epsilon_i}.$$

Proof. Since X_0 has positive probability of moving (and hence positive probability of getting larger) at each step, τ is finite with probability 1. Put

$$Y_t = X_t 1_{t \leq \tau} + n 1_{t > \tau}.$$

Let n_i be the number of times $Y_t = i$, where $0 < t \leq \tau$. Then $\tau = \sum_{i=0}^{n-1} n_i$, and $E[\tau] = \sum_{i=0}^{n-1} E[n_i]$. For $i < n$, $n_i = \sum_{t=0}^{\infty} 1_{Y_t = i}$ so

$$\begin{aligned}
E[n_i] &= \sum_{t=0}^{\infty} P(Y_t = i) \\
&\leq 1 + \sum_{t=1}^{\infty} P(Y_{t-1} = i-1, Y_t = i) + P(Y_{t-1} = i, Y_t = i) \\
&\quad + P(Y_{t-1} = i+1, Y_t = i) \\
&= 1 + \sum_{t=1}^{\infty} E[1_{Y_{t-1} = i-1, Y_t = i}] + P(Y_{t-1} = i) P(Y_t = i | Y_{t-1} = i) \\
&\quad + E[1_{Y_{t-1} = i+1, Y_t = i}] \\
&= 1 + E\big[U(i-1, i)\big] + E\big[D(i+1, i)\big] \\
&\quad + \sum_{t=1}^{\infty} P(Y_{t-1} = i) P(Y_t = i | Y_{t-1} = i)
\end{aligned}$$

where $D(i+1,i)$ is the number of downcrossing of X_ts from $i+1$ to i from time 0 to τ. Note that since we started at 0 and ended up at n, $D(i,i-1) = U(i-1,i)-1$, and so

$$\begin{aligned} E[n_i] &\leq 1 + E[U(i-1,i)] + E[U(i,i+1)] - 1 \\ &\quad + \sum_{t=1}^{\infty} P(Y_{t-1} = i) P(X_t = i | X_{t-1} = i) \\ &\leq n - (i-1) + n - i + \sum_{t=1}^{\infty} P(X_{t-1} = i)(1-\epsilon_i) \\ &= 2n - 2i + 1 + (1-\epsilon_i) \sum_{t=1}^{\infty} P(Y_{t-1} = i) \\ &\leq 2n - 2i + 1 + (1-\epsilon_i) E[n_i], \end{aligned}$$

so that

$$E[n_i] = \frac{2n-2i+1}{\epsilon_i}.$$

Summing over all i completes the proof of the corollary. □

We are now ready to prove Theorem 5.1.

Proof. At any time t, there will be a variable j for which the largest number of $x_{i,t}$ can be written in terms of $x_{j,T}$. Let A_t be the set of nodes with value either $x_{j,T}$ or $1 - x_{j,T}$. Clearly $|A_t| \geq 1$ for all time.

If $|A_t| = n$, we are done, so our goal will be to show that on average $|A_t|$ is growing. This can happen in two ways. First, a new node i might at time $t+1$ have value based on $x_{j,T}$, or a new node j' might arise such that more nodes are written in terms of $x_{j',T}$ than $x_{j,T}$. We only use this second effect to ensure that $|A_t|$ never falls below 1, otherwise we concentrate on the first means of growth (actually, at the first step it will move to 2 and always be at least 2 thereafter). Let I denote the random variable that picks a node uniformly from V; this node will then change color to the opposite of one of its (uniformly at random) chosen neighbors. Thus we have, for any node v,

$$P\big(|A_{t+1}| = |A_t| + 1 | \mathcal{F}_t, I = v\big) = 1 - P\big(|A_{t+1}| = |A_t| - 1 | \mathcal{F}_t, I = v\big).$$

Let $D_t = V \setminus A_T$, and set $d(v)$ to be the number of neighbors of v which are in D_t. Then

$$P\big(|A_{t+1}| = |A_t| - 1 | \mathcal{F}_t, I = v\big) \geq d(v)/r.$$

We now show that $|A_t|$ is a submartingale with respect to the canonical filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t = \sigma(A_s, s \leq t)$.

$$\begin{aligned} E\big[|A_{t+1}| - |A_t| \big| |A_t|\big] &\geq \frac{1}{n-1} \sum_{v \in D_t} P\big(|A_{t+1}| = |A_t| + 1 | \mathcal{F}_t, I = v\big) \\ &\quad - \sum_{v \in A_t} P\big(|A_{t+1}| = |A_t| - 1 | \mathcal{F}_t, I = v\big) \\ &= \frac{|D_t|}{n-1} - \frac{1}{n-1} \sum_v d(v)/r \\ &= \frac{|D_t|}{n-1} - \frac{1}{n-1} \frac{|D_t| r}{r} \\ &= 0. \end{aligned}$$

Hence $\{|A_t|, t = 1, 2, \ldots\}$ is a submartingale, and we may apply our corollary to the Upcrossing Inequality. We need to bound below $P(|A_{t+1}| \neq |A_t|)$ which is at least $P(|A_{t+1}| > |A_t|)$.

The minimum cut in the graph is c, and so at least c edges lie between nodes in A_t and D_t. Picking a node and neighbor at random is the same as picking an edge at random, followed by a random direction for that edge. Combining this with the fact that there are $nr/2$ edges gives us

$$P(|A_{t+1}| \neq |A_t|) \geq \frac{1}{2}\frac{c}{nr/2}.$$

Summing (note that we do not need the $i = 0$ term) gives $E[T] \leq n^3 r/c$. □

Note that, if the graph is such that any vertex can be mapped to any other vertex by an isomorphism, we have $P(|A_{t+1}| \neq |A_t|) \geq \frac{c}{nr/2}$, and $E[T] \leq n^3 \frac{r}{2c}$.

Aside from the fact of complete coupling, this shows that our modified chain also mixes in time $O(n^3 r/c)$. For the original chain where we do not flip colors, the mixing time is the same as our modified chain, but more work is needed to show it. We will not show how to create a complete coupling for the original chain (although it can be done), but simply a coupling.

Theorem 5.2. *The expected coupling time of the antivoter model chain is bounded above by* $n^3 r/c + nr(2n+1)$*, where* r *is the degree of the graph and* c *is the minimum cut.*

Proof. Suppose that we have two particles, X and Y on the chain. Let A_t be the set of nodes where they are colored the same way, and D_t be the set of nodes where they are colored differently. Then if $|A_t| \geq 1$, everything we proved in Theorem 5.1 is still true. That is, $|A_t|$ is still a submartingale and so our submartingale theorems still apply. The only change is in $P(|A_{t+1}| \neq |A_t|)$. In the case that $|A_t| = 0$, this probability may be lower.

Often, when constructing a coupling chain, we desire to update the same sites of the graph for all particles. When $|A_t| = 0$, however, this strategy is unproductive. The fact that $|A_t| = 0$ means that X and Y are colored the opposite of one another at every site. If X and Y are updated at the same site with the same neighbor, then they will continue to be colored the opposite of one another everywhere for all time. Hence when $|A_t| = 0$, we choose nodes and neighbors for X and Y to color independently of each other.

Suppose we look at the cut defined in X (or Y) by the two coloring. The number of edges which are colored the same way at both endpoints is just the total number of edges minus the size of this cut. If C is the size of the maximum unweighted cut in the graph, then $nr/2 - C$ edges are colored the same way. This value is positive because the graph is not bipartite, and hence the maximum cut does not include all of the edges of the graph.

Instead of picking first a node then a neighbor, we envision the antivoter model as picking a random edge, then a random direction for that edge.

The head of the edge gets changed to the opposite color of the tail. Now if any of the $nr/2 - C$ edges which are colored the same way at tail and head are picked, the color of a node will be changed in X. Since we are in a nonbipartite graph, $C < nr/2$, and this probability is at least $1/(nr/2)$. When X picks this edge, Y does not, and vice versa so that the probability that the two share at least one node after one step is $4/(nr)$.

Therefore by our corollary, the expected time spent at $|A_t| = 0$ is at most $(nr/4)(2n+1)$. All of the expected values of n_i for $i > 0$ remain the same, and so we find that

$$E[\tau] \leq \frac{nr(2n+1)}{2} + \sum_{i=1}^{n-1} \frac{nr(2n-2i+1)}{2c} = \frac{nr(2n+1)}{2} + \frac{n^3 r}{2c}.$$

Because $c < n$, this last expression is $O(n^3 r/c)$. □

5.3. The complete graph

As before, if $x \in \{0,1\}^V$ is a coloring, we let $W(x) = \sum_{v \in V} x(v)$. The complete graph has the unique property that the statistic W determines the coloring up to a permutation of the nodes. This enables us to compute the stationary distribution of the complete graph directly. The recipe for computing is also given in Aldous [2].

Theorem 5.3. *Consider the antivoter model on the complete graph. Then*

$$\pi(x) = \frac{\binom{n-2}{W(x)-1}}{\binom{2n-2}{n-1}},$$

and

$$\pi(\{x : W(x) = i\}) = \frac{\binom{n}{i}\binom{n-2}{i-1}}{\binom{2n-2}{n-1}}.$$

Proof. Let X_t be a particle run on the antivoter model on the complete graph, and let $W_t = W(X_t)$. Let A_t be the set of nodes colored 0 at time t, and B_t be the set of nodes colored 1 (so $W_t = |B_t|$). Then $P(W_{t+1} - W_t = 0)$ is the probability that an edge between A_t and B_t is chosen. There are $|A_t \| B_t| = W_t(n - W_t)$ such edges. $P(W_{t+1} - W_t = 1)$ is just the probability that an edge with both endpoints in A_t is chosen, and $W_{t+1} - W_t = -1$ occurs when both endpoints of the chosen edge lie in B_t. Hence $P(W_{t+1} - W_t = 1) = \binom{W_t}{2}/\binom{n}{2}$ and $P(W_{t+1} - W_t = -1) = \binom{n-W_t}{2}/\binom{n}{2}$. Therefore $P(W_{t+1} = a | W_t, W_{t-1}, \ldots) = P(W_{t+1} = a | W_t)$ and W_t is a Markov chain on $\{0, \ldots, n\}$. Since it is a birth death chain as well, it is reversible, so

$$\pi_W(i-1)P(W_t = i-1, W_{t+1} = i) = \pi_W(i)P(W_t = i, W_{t+1} = i-1).$$

After canceling the $\binom{n}{2}$ terms that arise in the probabilities, we have

$$\begin{aligned} \frac{\pi_W(i)}{\pi_W(i-1)} &= \frac{\binom{n-(i-1)}{2}}{\binom{i}{2}} \\ &= \frac{n-(i-1)}{i} \cdot \frac{n-i}{i-1} \end{aligned}$$

Note that for all t, $1 \leq W_t \leq n-1$. Therefore we have that for $n - 1 \geq i > 1$

$$\begin{aligned} \frac{\pi_W(i)}{\pi_W(1)} &= \frac{(n-1)(n-2)\cdots(n-(i-1))}{i!} \frac{(n-2)(n-3)\cdots(n-i)}{(i-1)!} \\ &= \frac{(n-1)!}{i!(n-i)!} \cdot \frac{(n-2)!}{(i-1)!(n-(i+1))!} \\ &= \frac{1}{n}\binom{n}{i}\binom{n-2}{i-1} \end{aligned}$$

Set $z(i) = \binom{n}{n-i}\binom{n-2}{i-1}$, so that $z(i) = n\pi_W(i)/\pi_W(1)$. Then $\sum_{i=0}^{n-1} z(i) = n/\pi_W(1)$. Examining this sum closer, we see that it is just the number of ways of choosing $n-1$ objects from an initial set of size $2n-2$. Hence $\pi_W(1) = n/\binom{2n-2}{n-1}$, and

$$\pi_W(i) = \frac{\binom{n}{i}\binom{n-2}{i-1}}{\binom{2n-2}{n-1}}.$$

Finally, consider that each of $\binom{n}{i}$ colorings have $W(x) = i$. Hence each coloring has $\pi(x) = \pi_W(W(x))/\binom{n}{i}$ and

$$\pi(x) = \frac{\binom{n-2}{W(x)-1}}{\binom{2n-2}{n-1}}.$$

□

Intuitively, we might understand this stationary distribution as follows. From the conditions of the antivoter model, all possible configurations must have at least one node colored with the first color, and at least one node colored with the second color, if the system is in stationarity. Secondly observe that by symmetry all configurations with the same number of vertices colored with the first color are equally likely; $W(x)$ is a sufficient statistic. Let $W(X)$ be that count for a random configuration X in stationarity. Then, conditioning on having at least one node of each color yields that, for each i, $P(W(X) = i)$ must be proportional to $\binom{n-2}{i-1}$. Suppose we have an underlying process where the vertices are colored independently, choosing each of the two colors with probability 1/2.

Then the expected number of configurations from this independent chain that are hit by $W(X)$ is proportional to $\binom{n}{i}\binom{n-2}{i-1}$. From this it is intuitively clear that $\pi(x)$ is proportional to $\binom{n}{W(x)}\binom{n-2}{W(x)-1}$.

We can also say something more about the mixing time of the antivoter chain on the complete graph. Since the minimum cut in the complete graph is of size $n-1$, the mixing time for this chain will be $O(n^3)$.

5.4. The multipartite graph

Let $\mathcal{M}(k,l)$ denote the multipartite graph with k groups of l vertices each. Thus there are $n = kl$ vertices in total. Each node within a group is connected to all vertices in all the other groups, but to none of the vertices in the same group. Thus each node has degree $r = (k-1)l = n-l$, and there are $rn = n(n-l)$ edges total in the graph. The complete graph corresponds to the case $k = n, l = 1$. Intuitively, if l is not too large, the graph $\mathcal{M}(k,l)$ resembles the complete graph $\mathcal{M}(n,1)$, and thus the stationary distribution on $\mathcal{M}(k,l)$ should be close to the one for the complete graph, described above.

Moreover, using Stein's method, Rinott and Rotar [12] proved the following normal approximation for the number $W = W(\mathbf{X})$ of ones in the Markov chain $\mathbf{X}$ when started in stationarity.

Theorem 5.4. *For any $k \geq 3$, there are constants $C > 0$ and $\bar{\sigma} > 0$ such that $VarW \geq \bar{\sigma}^2 n$ and*

$$\sup_x \left| P\left(\frac{W}{\bar{\sigma}\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq \frac{C}{\bar{\sigma}^2\sqrt{n}}\left(\frac{1}{\bar{\sigma}} + \sqrt{\frac{k}{k-1}}\right).$$

From this it follows that the count $W(X)$ for the antivoter model on the complete graph satisfies the Central Limit Theorem (CLT), and thus so will the count on the multipartite graph $\mathcal{M}(k,l)$ provided l is not too large. However, if the groups contain only a small number of vertices each, then an approximation of stationary distribution of the count on the multipartite graph with the stationary distribution of the count on the complete graph will be more accurate. This section is devoted to making the above heuristic precise. Note that for the bipartite graph $\mathcal{M}(2,n/2)$ the CLT will not hold; the stationary distribution is trivial (see Rinott and Rotar [12]). Thus for large l we expect the approximation to break down. The main tool for assessing the proximity of the stationary distributions is Stein's method, applied as described in Holmes [6]. As observed by Rinott and Rotar [12], the transition probabilities of the Markov chain describing the antivoter process, when restricted to $W(\mathbf{X})$, can be derived as follows. Set

$$\begin{aligned} a(\mathbf{X}) &= \text{the number of edges } (i,j) \quad \text{with } X_i = X_j = 1, \\ b(\mathbf{X}) &= \text{the number of edges } (i,j) \quad \text{with } X_i = X_j = -1, \\ c(\mathbf{X}) &= \text{the number of edges } (i,j) \quad \text{with } X_i \neq X_j. \end{aligned}$$

Observe that

$$W(\mathbf{X}) = \frac{2a(\mathbf{X}) + c(\mathbf{X})}{r},$$

and thus

$$n - W(\mathbf{X}) = \frac{2b(\mathbf{X}) + c(\mathbf{X})}{r}. \tag{5.1}$$

Then we have for $W = W(\mathbf{X}(t)), W' = W(\mathbf{X}(t+1))$, which Rinott and Rotar [12] prove to be an exchangeable pair, that

$$P(W' - W = -1 \,|\, \mathbf{X}) = 2a(\mathbf{X})/rn, \qquad P(W' - W = 1 \,|\, \mathbf{X}) = 2b(\mathbf{X})/rn. \tag{5.2}$$

The above process can be viewed as a birth and death chain with birth rate $\beta_{l,k}$ satisfying

$$E(\beta_{l,W}|\mathbf{X}) = 2b(\mathbf{X})/rn$$

and with death rate $\delta_{l,k}$ satisfying

$$E(\delta_{l,W}|\mathbf{X}) = 2a(\mathbf{X})/rn;$$

the index l refers to the decomposition in the multipartite graph. The corresponding generator for this chain, using the scaling in Holmes [6] with $1/\lambda$, where $\lambda = 2/n$ following [12], is

$$T_l f(k) = n\big(\beta_{l,k} f(k+1) - \delta_{l,k} f(k)\big).$$

In the complete graph the situation is particularly easy: if W vertices are labelled with the first color, we have

$$a(\mathbf{X}) = \frac{\binom{W}{2}}{\binom{n}{2}}$$

and

$$b(\mathbf{X}) = \frac{\binom{n-W}{2}}{\binom{n}{2}}.$$

Thus the birth and death rates are

$$\beta_k = \beta_{1,k} = \frac{\binom{n-k}{2}}{\binom{n}{2}}$$

and

$$\delta_k = \delta_{1,k} = \frac{\binom{k}{2}}{\binom{n}{2}}.$$

The corresponding generator is

$$T_1 f(k) = n\left(\frac{\binom{n-k}{2}}{\binom{n}{2}} f(k+1) - \frac{\binom{k}{2}}{\binom{n}{2}} f(k)\right).$$

Let π_l be the stationary distribution for W on $\mathcal{M}(k,l)$, and let π be the stationary distribution for W on the complete graph. Then, following Holmes [6], we have for all w that

$$\big|\pi_l(w) - \pi(w)\big| = \big|E(T_l - T_1)f(W)\big|,$$

where f is the pseudo-inverse of $h(w) = \mathbf{1}_k(w)$, that is, f solves the Stein equation for h

$$\mathbf{1}_k(w) - \pi(k) = T_1 f(w). \tag{5.3}$$

Thus it suffices to bound $|E(T_l - T_1)f(W)|$, for all such functions f. First we will derive a bound on $\|f\| = \sup_x |f(x)|$.

Lemma 5.2. *Let $n \geq 6$. For each $k = 0, 1, \ldots, n$ there is a function f that solves the Stein equation* (5.3) *for the complete graph such that*

$$\|f\| \leq \frac{7}{n}.$$

Proof. The existence is shown in Holmes [6]. As, for the complete graph, it is easy to check that the birth rates β_k are strictly decreasing in k, whereas the death rates δ_k are strictly increasing in k, from Holmes [6] we have

$$\big\|f(w)\big\| \leq \Delta U 1_k(k),$$

where

$$\Delta U 1_k(k) = \frac{1}{n}\pi(k)\left(\frac{1 - S_k}{\beta_k \pi(k)} + \frac{S_{k-1}}{\beta_{k-1}\pi(k-1)}\right),$$

with

$$S_k = \sum_{j=0}^{k} \pi(j).$$

Moreover we know that

$$\pi(k) = \frac{\beta_{k-1}}{\delta_k}\frac{\beta_{k-2}}{\delta_{k-1}} \cdots \frac{\beta_1}{\delta_2}\pi(1).$$

Thus

$$\begin{aligned}\Delta U1_k(k) &= \frac{1}{n}\left(\frac{1-S_k}{\beta_k}+\frac{S_{k-1}\pi(k)}{\beta_{k-1}\pi(k-1)}\right)\\ &= \frac{1}{n}\left(\frac{1-S_k}{\beta_k}+\frac{S_{k-1}}{\delta_k}\right)\\ &< \frac{1}{n}\left(\frac{1-S_k}{\beta_k}+\frac{S_k}{\delta_k}\right),\end{aligned}$$

where we bounded S_{k-1} by S_k in the last step. It is easy to see that the last expression is maximized for $\beta_k = \delta_k$, which translates to $k = \frac{n}{2}$ for n even, and $k = \frac{n-1}{2}$ or $k = \frac{n+1}{2}$ for n odd. Let us assume first that n is even. We have

$$\begin{aligned}\Delta U1_k(k) &< \frac{1}{n}\left(\frac{1-S_{n/2}}{\beta_{n/2}}+\frac{S_{n/2}}{\delta_{n/2}}\right)\\ &= \frac{1}{n}\frac{\binom{n}{2}}{\binom{n/2}{2}}\\ &= 4\frac{n-1}{n(n-2)}\\ &< \frac{7}{n},\end{aligned}$$

as $n \geq 6$. For n odd we have

$$\begin{aligned}\Delta U1_k(k) &< \frac{1}{n}\frac{\binom{n}{2}}{\binom{(n-1)/2}{2}}\\ &= 4\frac{n}{n(n-3)}\\ &\leq \frac{7}{n},\end{aligned}$$

as $n \geq 7$. □

Note that the case $n < 6$, which is not covered by the above result, is not of interest of us, since for those cases the only possible non-bipartite but multipartite graph is the complete graph itself. Now we have the ingredients to prove an approximation result.

Theorem 5.5. *Let π_l be the stationary distribution for W on $\mathcal{M}(k,l)$, let $n = kl$, and let π be the stationary distribution for W on the complete graph on n nodes. Then*

$$\|\pi_l - \pi\|_{TV} \leq \frac{7}{n}\left(\frac{2l(n-l)+(l-1)n^2}{(n-1)(n-l)}\right).$$

From Theorem 5.4 it follows in particular that π_l satisfies the CLT if l is not too large. Moreover note that Theorem 5.5 gives a bound for any finite l, n; we are not required to describe a limiting procedure for the different graphs. Note also that, for large n, the bound in Theorem 5.5 is of the order $\frac{1}{k-1}$, so that for large k and small l this approximation with the stationary distribution on the complete graph might indeed be more accurate than the normal approximation. However, both here and in Theorem 5.4 only upper bounds are given.

Proof. From the above, it suffices to bound $E(T_l - T_1)f(W)$ for all f which are solution of the Stein equation (5.3) for some indicator function h. We have

$$E(T_l - T_1)f(W) = nE\big((\beta_W - \beta_{l,W})f(W+1) - (\delta_W - \delta_{l,W})f(W)\big).$$

For the first difference on the right-hand side, note that from (5.1), $2b(\mathbf{X}) = r(n - W) - c(\mathbf{X})$. Thus

$$\begin{aligned}
&E(\beta_W - \beta_{l,W})f(W+1)\\
&\quad = E(\beta_W - E\{\beta_{l,W}|\mathbf{X}\})f(W+1)\\
&\quad = E\bigg(\frac{(n-W)(n-W-1)}{n(n-1)} - \frac{2b(\mathbf{X})}{n(n-l)}\bigg)f(W+1)\\
&\quad = E\bigg(\frac{(n-W)(n-W-1)}{n(n-1)} - \frac{(n-l)(n-W)+c(\mathbf{X})}{n(n-l)}\bigg)f(W+1)\\
&\quad = E\bigg(\frac{(n-W)W}{n(n-1)} - \frac{c(\mathbf{X})}{n(n-l)}\bigg)f(W+1).
\end{aligned}$$

Similarly we obtain

$$\begin{aligned}
E(\delta_W - \delta_{l,W})f(W) &= E\bigg(\frac{W(W-1)}{n(n-1)} - \frac{2a(\mathbf{X})}{n(n-l)}\bigg)f(W)\\
&= -E\bigg(\frac{(n-W)W}{n(n-1)} - \frac{c(\mathbf{X})}{n(n-l)}\bigg)f(W).
\end{aligned}$$

Hence

$$\begin{aligned}
&\big|E(T_l - T_1)f(W)\big|\\
&\quad = \bigg|nE\bigg(\frac{(n-W)W}{n(n-1)} - \frac{c(\mathbf{X})}{n(n-l)}\bigg)(f(W+1)+f(W))\bigg|\\
&\quad \le 2\|f\|\frac{1}{(n-1)(n-l)}E\big|(n-l)W(n-W) - (n-1)c(\mathbf{X})\big|\\
&\quad = 2\|f\|\frac{1}{(n-1)(n-l)}E\big|(n-l)(W(n-W) - c(\mathbf{X})) + (l-1)c(\mathbf{X})\big|.
\end{aligned}$$

If there are W nodes labelled 1, then at most l of them can be in the same component. Also there will then be $n - W$ nodes labelled 0, and again at most l of them can be in the same component. Thus

$$c(\mathbf{X}) \ge (W-l)(n-W-l)$$

and

$$\begin{aligned}
W(n-W) - c(\mathbf{X}) &\le W(n-W) - (W-l)(n-W-l)\\
&= l(n-l).
\end{aligned}$$

Together with the observation that $c(\mathbf{X}) \le W(n-W)$, this yields

$$\begin{aligned}
\big|E(T_l - T_1)f(W)\big| &\le 2\|f\|\frac{1}{(n-1)(n-l)}\big(l(n-l) + (l-1)nEW\big)\\
&= 2\|f\|\frac{1}{(n-1)(n-l)}\bigg(l(n-l) + (l-1)\frac{n^2}{2}\bigg).
\end{aligned}$$

Lemma 5.2 now gives the assertion. □

5.5. The empirical Stein's method

The empirical Stein method is described in detail in Stein [13]. In the context of the antivoter model it can be sketched as follows.

As exact sampling from the stationary distribution is rather time-consuming, in particular in the tails of the distribution. To estimate the stationary distribution we draw one exact sample and let the chain run for a long time with this initial value. The classical frequency estimate of the stationary distribution π of W is

$$\hat{\pi}(w) = \frac{1}{T} \sum_{t=1}^{T} 1_{W(t)=w}.$$

Here $W(t)$ denotes the number of 1's at time t, where the antivoter chain has been started at a configuration drawn from its stationary distribution. However, this frequency estimate can be improved considerably by making use of the conditional distributions in the problem.

From the above, we already know that a central limit theorem holds for the complete graph, as well as for suitable multipartite graphs. Thus we expect the stationary distribution π to be roughly normal, for the antivoter model on multipartite graphs. Therefore this example is an interesting test case for the empirical Stein method, and it will illustrate its power.

Let $(W, W') = (W(t), W(t+1))$ denote two consecutive instances of the number of ones in the antivoter model, started in stationarity. Rinott and Rotar [12] have shown that this is an exchangeable pair. We have already made use of the property (5.2)

$$P(W' - W = -1 \,|\, \mathbf{X}) = 2a(\mathbf{X})/rn, \qquad P(W' - W = 1 \,|\, \mathbf{X}) = 2b(\mathbf{X})/rn.$$

This gives another estimate of the ratios of the stationary distribution

$$\frac{\tilde{\pi}(x+1)}{\tilde{\pi}(x)} = \frac{\widetilde{P}(W' = w+1 | W = w)}{\widetilde{P}(W' = w | W = w+1)}.$$

Now we estimate

$$\widetilde{P}(W' = w+1 | W = w) = \frac{2 \sum_{t=1}^{t} 1_{W(t)=w} a(\mathbf{X}(t))/rn}{\sum_{t=1}^{t} 1_{W(t)=w}}$$

and

$$\widetilde{P}(W' = w-1 | W = w) = \frac{2 \sum_{t=1}^{t} 1_{W(t)=w} b(\mathbf{X}(t))/rn}{\sum_{t=1}^{t} 1_{W(t)=w}}.$$

Using the fact that probabilities sum up to 1 thus leads to a new estimate $\tilde{\pi}$, which we call the "empirical Stein estimate", for the stationary distribution π.

Keeping track of the whole configuration $\mathbf{X}(t)$ contains more information than just keeping track of the sums $W(t)$, and hence it is not surprising that this estimate is more accurate in practice than the standard frequency estimate. Figures 1–6 show quantile-quantile and distribution plots for two different tripartite graphs. First we consider the tripartite graph on 300 nodes. The estimated distribution should be approximately normal, so we compare both estimates with the corresponding standardized normal distribution, using the estimated standard deviations (Figure 1).

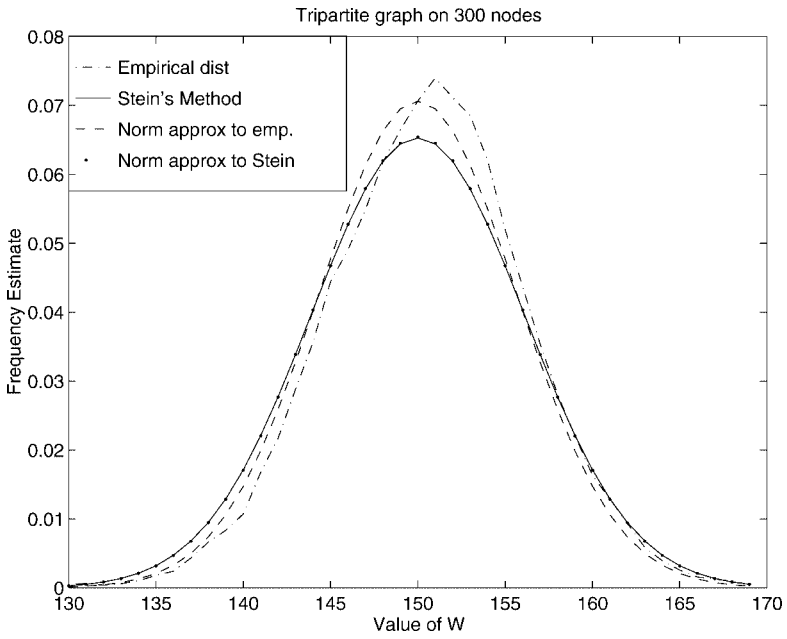

Figure 1: Distribution plot for W

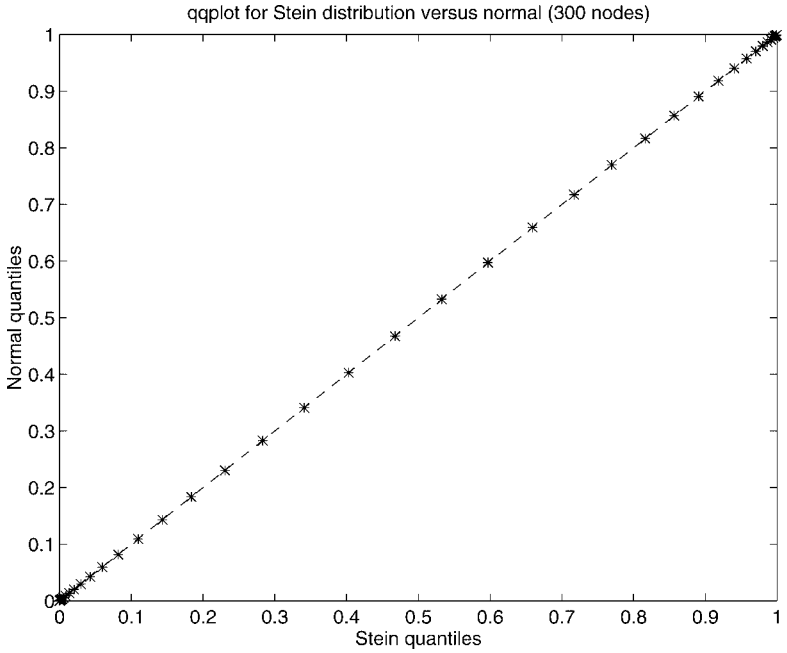

Figure 2: Quantile-quantile plot for W, estimated by empirical Stein

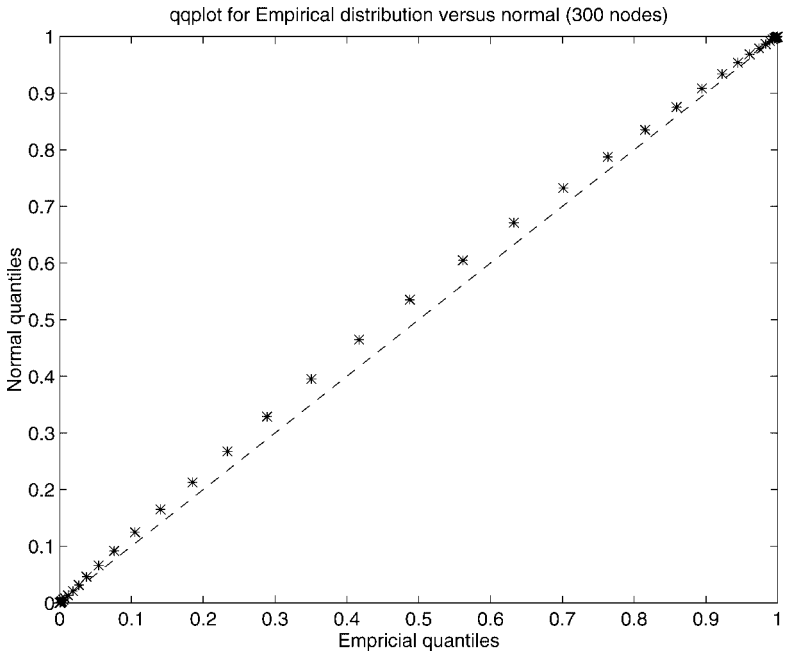

Figure 3: Quantile-quantile plot for W, estimated by empirical frequency method

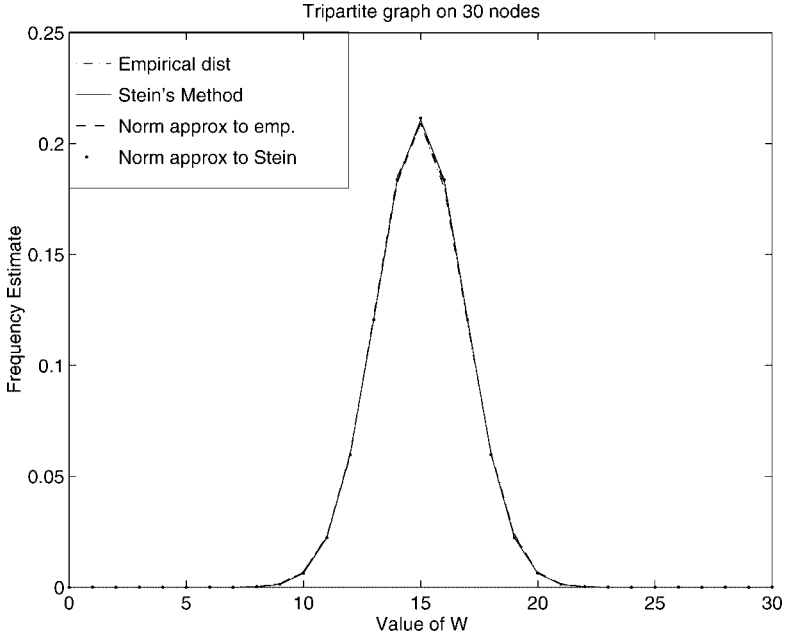

Figure 4: Distribution plot for W

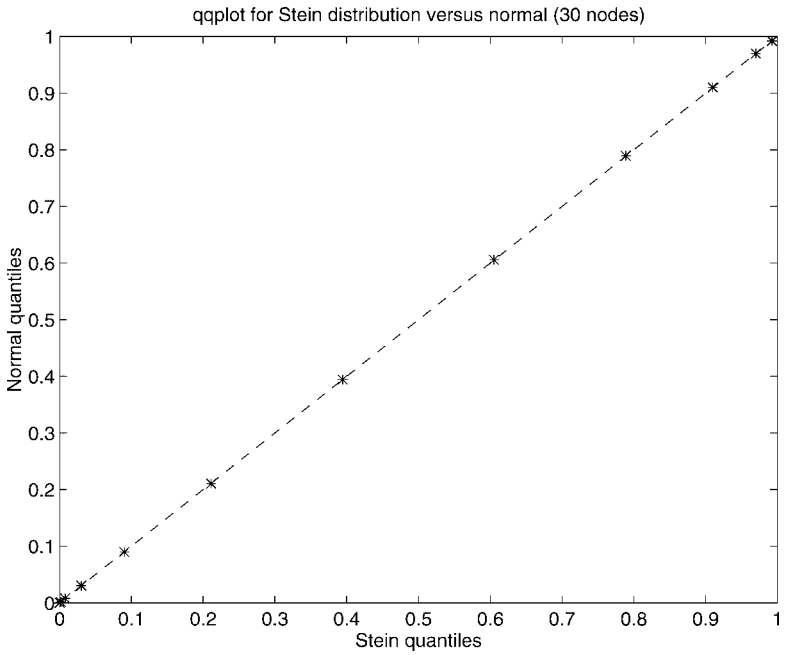

Figure 5: Quantile-quantile plot for W, estimated by empirical Stein

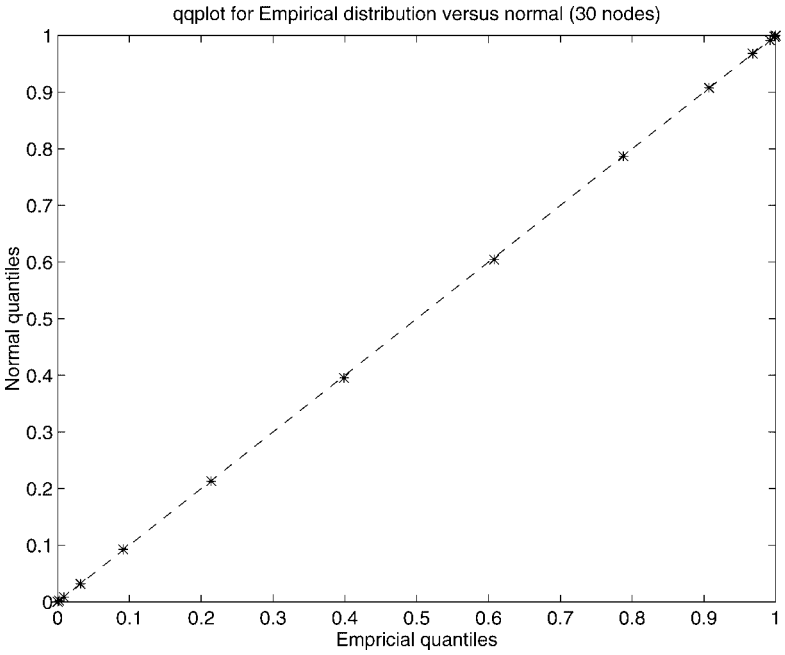

Figure 6: Quantile-quantile plot for W, estimated by empirical frequency method

The empirical Stein estimate is much closer to its normal analog than the frequency estimate is to its, after a moderate run length of 100,000 runs. The quantile-quantile plots for the frequency estimate and for the empirical Stein estimate illustrate again that the empirical Stein estimate (Figure 3) is much closer to the normal than the frequency estimate (Figure 2); moreover it is considerably smoother. The same is true for the tripartite graph on 30 nodes. Here a run length of 100,000 puts both estimates much closer to the corresponding normal distribution (Figure 4); yet the quantile-quantile plots reveal that the empirical Stein estimate is even closer to normality than the frequency estimate is (Figures 5 and 6).

The empirical Stein method outperforms the standard frequency estimate also in all other multipartite graphs we have looked at. This poses intriguing open questions - why is this estimate so much better? Would the same be true for the antivoter model on other graphs? More examples remain to be considered.

Acknowledgments

We would like to thank Charles Stein and Richard Arratia for many helpful comments. Also our thanks to Persi Diaconis and Susan Holmes, who organized the wonderful summer month meeting in Stanford where this work started. Morevoer we would like to acknowledge Peter Donnelly for pointing out the connection with the Moran model. This work was partially supported by an Office of Naval Research Fellowship and NSF grants CCR-9307391, CCR-9700029, and DMS-9505155, and ONR grant N00014-96-1-00500 for MH and by an EPSRC grant no. GT/R52183/01 to GR.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In A. Dold and B. Eckmann, editors, *Séminaire de Probabilités XVII 1981/1982*, volume 986 of *Springer-Verlag Lecture Notes in Mathematics*, 243–297. Springer-Verlag, New York, 1983. MR770418

[2] D. Aldous and J.A. Fill. *Reversible Markov Chains and Random Walks on Graphs.* Monograph in preparation, 2003. `http://www.stat.berkeley.edu/~aldous/RWG/book.html`

[3] W. Doeblin. Exposé de la théorie des chains simples constantes de Markov à un nombre fini d'états. Rev. Math. de l'Union Interbalkanique, volume 2, 77–105, 1993.

[4] P. Donnelly. The transient behaviour of the Moran model in population genetics. *Math. Proc. Camb. Phil. Soc.* 95:349–358, 1984. MR735377

[5] P. Donnelly and D. Welsh. The antivoter problem: random 2-colourings of graphs. *Graph Theory and Combinatorics*, 133–144, 1984. MR777170

[6] S. Holmes. Stein's method for birth and death chains. In this volume, *Expository Lectures and Applications of Stein's method*, Chapter 3, 2004.

[7] M. L. Huber. Exact sampling and approximate counting techniques. In *Proceedings of the 30th ACM Symposium on the Theory of Computing*, 1998.

[8] T. Liggett. *Interacting Particle Systems.* Springer, New York, 1985. MR776231

[9] N. Matloff. Ergodicity conditions for a dissonant voting model. *Annals of Probability*, 5: 371–386, 1977. MR445646

[10] P.A.P. Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.* 54: 60–71, 1958. MR127989

[11] S. C. Port. *Theoretical Probability for Applications*. Wiley-Interscience, 1994. MR1265493

[12] Y. Rinott and V. Rotar. On coupling constructions and rates in the clt for dependent summands with applications to the anti-voter model and weighted U-statistics. *Ann. Appl. Probab.* 7: 1080–1105, 1997. MR1484798

[13] C. Stein, P. Diaconis, S. Holmes and G. Reinert. Use of exchangeable pairs in the analysis of simulations. In this volume, *Stein's Method: Expository Lectures and Applications*, Chapter 1, 2004.