

The Voting Power Gap: Identifying Racial Gerrymandering with a Discrete Voter Model

A THESIS PRESENTED
BY
HAKEEM OLAKUNLE ISA ANGULU
TO
THE DEPARTMENTS OF COMPUTER SCIENCE AND STATISTICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF ARTS
IN THE SUBJECTS OF
COMPUTER SCIENCE AND STATISTICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2020

©2020 – HAKEEM OLAKUNLE ISA ANGULU
ALL RIGHTS RESERVED.

The Voting Power Gap: Identifying Racial Gerrymandering with a Discrete Voter Model

ABSTRACT

Section 2 of the Voting Rights Act of 1965 (VRA)⁶ prohibits voting practices or procedures that discriminate based on race, color, or membership in a language minority group, and is often cited by plaintiffs seeking to challenge racially-gerrymandered districts in court.

In 1986, with *Thornburg v. Gingles*⁵, the Supreme Court held that in order for a plaintiff to prevail on a section 2 claim, they must show that:

1. the racial or language minority group is sufficiently numerous and compact to form a majority in a single-member district
2. that group is politically cohesive
3. and the majority votes sufficiently as a bloc to enable it to defeat the minority's preferred candidate

All three conditions are notoriously hard to show, given the lack of data on how people vote by race.

In the 1990s and early 2000s, Professor Gary King's ecological inference method tackled the second condition: racially polarized voting, or racial political cohesion. His technique became the standard

technique for analyzing racial polarization in elections by inferring individual behavior from group-level data. However, for more than 2 racial groups or candidates, that method hits computational bottlenecks.

A new method of solving the ecological inference problem, using a mixture of contemporary statistical computing techniques, is demonstrated here. It can be used for multiple racial groups and candidates, and is shown to work well on randomly-generated mock election data and case studies in Chicago. This method, the Discrete Voter Model, is abstract and extensible to many ecological inference problems, even those unrelated to voting.

Contents

o INTRODUCTION	I
1 BACKGROUND AND LITERATURE REVIEW	6
1.1 Racial Gerrymandering and the Courts	6
1.2 Ecological Inference	II
1.3 Discretization	15
2 METHODS	22
2.1 phc	23
2.2 expec_votes	29
2.3 prob_votes	31
2.4 elect	33
2.5 dvm	35
2.6 Evaluation	38
3 RESULTS	40
3.1 Experiment Specification	41
3.2 Experiment Results	43
4 DISCUSSION	55
4.1 The Effect of DVM's Parameters	57
4.2 Case Studies: Chicago 2015 and 2019 Mayoral Runoffs	61
4.3 Implications	71
4.4 Future Work	72
5 CONCLUSION	73
APPENDIX A APPENDIX	75
A.1 Bar Plots Comparing Experiment Results For All Elections	80
REFERENCES	97

List of Tables

1.1	A 3×3 Table of Voting By Race, using Counts of Voting Age Population	12
1.2	Quantities in King's EI	13
1.3	Parameters in King's EI	14
2.1	Partitions of 10 votes among 3 demographic groups	32
3.1	Non-Election Variables for DVM Experiments	41
3.2	Demographic Distributions Tested in the 2×2 Case	42
3.3	Demographic Voting Probabilities Tested in the 2×2 Case	42
3.4	Demographic Distributions Tested in the 3×2 Case	43
3.5	Demographic Voting Probabilities Tested in the 3×2 Case	43
3.6	Experiment Results for the RWM Kernel, Scoring by Expectation, for 2×2 Elections	45
3.7	Experiment Results for the RWM Kernel, Scoring by Probability, for 2×2 Elections	46
3.8	Experiment Results for the HMC Kernel for 2×2 Elections	47
3.9	Experiment Results for King's Ecological Inference Method on 2×2 Elections . .	47
3.10	Experiment Results for the RWM Kernel, Scoring by Expectation, for 3×2 Elections	48
3.11	Experiment Results for the HMC Kernel, Scoring by Expectation, for 3×2 Elections	48

List of Figures

1.1	The Creation of Texas' 25th Congressional District (orange) in 2003 ¹²	11
1.2	Kernel Density Plot of the Beta(0.1, 0.1) Distribution	16
1.3	A Bivariate Multimodal Distribution	17
2.1	Example of a Rank 3, Granularity 10 Probabilistic Hypercube	26
2.2	Example of a Rank 2, Granularity 10 Probabilistic Hypercube	27
2.3	The Probability Distribution within of a Rank 2, Granularity 10 Probabilistic Hypercube	28
3.1	Runtime of the Configurations on the $d_1 \times v_1$ Election	49
3.2	MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_1$ Election	50
3.3	MSE of the MLE of the Configurations' DVM on the $d_1 \times v_1$ Election	51
3.4	Runtime of the Configurations on the $d_4 \times v_4$ Election	52
3.5	MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_4$ Election	53
3.6	MSE of the MLE of the Configurations' DVM on the $d_4 \times v_4$ Election	54
4.1	Choropleth Maps of the Four Largest Racial Groups in Chicago. Figure 2 in MGGG's Chicago Report. ²²	63
4.2	A Trace of the Sigmoid Difference in Expectation of the Probabilistic Hypercubes in the DVM Chain for Chicago's 2015 Mayoral Runoff Election	65
4.3	A Trace of the Sigmoid Difference in Expectation of the Probabilistic Hypercubes in the DVM Chain for Chicago's 2019 Mayoral Runoff Election	66
4.4	The Mean PHC for the 2015 Chicago Mayoral Runoff, for Rahm Emanuel	67
4.5	The MLE for the 2015 Chicago Mayoral Runoff, for Rahm Emanuel	68
4.6	The Mean PHC for the 2019 Chicago Mayoral Runoff, for Lori Lightfoot	69
4.7	The MLE for the 2019 Chicago Mayoral Runoff, for Lori Lightfoot	70
A.1	Runtime of the Configurations on the $d_1 \times v_1$ Election	80
A.2	Runtime of the Configurations on the $d_1 \times v_2$ Election	81
A.3	Runtime of the Configurations on the $d_2 \times v_1$ Election	81
A.4	Runtime of the Configurations on the $d_2 \times v_2$ Election	82
A.5	Runtime of the Configurations on the $d_3 \times v_1$ Election	82
A.6	Runtime of the Configurations on the $d_3 \times v_2$ Election	83
A.7	MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_1$ Election	84
A.8	MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_2$ Election	85

A.9	MSE of the Mean PHC of the Configurations' DVM on the $d_2 \times v_1$ Election	85
A.10	MSE of the Mean PHC of the Configurations' DVM on the $d_2 \times v_2$ Election	86
A.11	MSE of the Mean PHC of the Configurations' DVM on the $d_3 \times v_1$ Election	86
A.12	MSE of the Mean PHC of the Configurations' DVM on the $d_3 \times v_2$ Election	87
A.13	Runtime of the Configurations on the $d_4 \times v_3$ Election	88
A.14	Runtime of the Configurations on the $d_4 \times v_4$ Election	89
A.15	Runtime of the Configurations on the $d_5 \times v_3$ Election	89
A.16	Runtime of the Configurations on the $d_5 \times v_4$ Election	90
A.17	Runtime of the Configurations on the $d_6 \times v_3$ Election	90
A.18	Runtime of the Configurations on the $d_6 \times v_4$ Election	91
A.19	MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_3$ Election	92
A.20	MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_4$ Election	93
A.21	MSE of the Mean PHC of the Configurations' DVM on the $d_5 \times v_3$ Election	93
A.22	MSE of the Mean PHC of the Configurations' DVM on the $d_5 \times v_4$ Election	94
A.23	MSE of the Mean PHC of the Configurations' DVM on the $d_6 \times v_3$ Election	94
A.24	MSE of the Mean PHC of the Configurations' DVM on the $d_6 \times v_4$ Election	95

THIS THESIS IS DEDICATED TO MY MOM, MY FIRST EXAMPLE AND SOURCE OF STRENGTH.
MAY THIS *DISCRETE VOTER MODEL* CONTRIBUTE TO JUSTICE AND MAKE HER PROUD.

Acknowledgments

This thesis is the culmination of four years of undergraduate studies and several months of research.

I would not have gotten to this point without the constant support of many individuals.

I would like to express my deep gratitude to Professor Moon Duchin, my thesis advisor, but also my mentor and teacher for the past two years. She inspired me, encouraged me, taught me, and believed in me, and without her this thesis would have been impossible.

I would like to thank my mom for her unwavering love and selflessness. She has given so much of herself to ensure that I was able to follow my dreams, and for that I am immensely thankful. Although I am far from her and home, her care propels me forward.

I would like to thank Dr. Thomas Weighill, whose patience, ingenuity, generosity, and brilliance contributed greatly to this research.

I would like to thank the researchers of the Metric Geometry and Gerrymandering Group, particularly Ruth Buck and Dr. Daryl DeFord. They were instrumental in my early understandings of computational research on gerrymandering, and gave me opportunities to grow.

I would like to thank Professor Latanya Sweeney and Jinyan Zang. They provided me with patient guidance, enthusiastic encouragement and useful feedback on this research.

Finally, I would like to thank my classmates and friends. They have toiled beside me at all hours, listened to me, helped me and held me throughout this experience. Thank you.

O

Introduction

THE VOTING RIGHTS ACT of 1965 (VRA)⁶ was the result of decades of activism and advocacy around unencumbered suffrage in the United States. Specifically, this act was meant to “enforce the fifteenth amendment to the constitution,” which was ratified 95 years prior, and came after African Americans in the South protested the tremendous obstacles to voting they encountered, including poll taxes,

literacy tests, harassment, intimidation, physical violence, and other systemic and infrastructural facets of the practical application of the right to vote. The United States Department of Justice has cited it as “the single most effective piece of civil rights legislation ever passed by Congress.”⁷

Section 2 of that act specifically prohibits any jurisdiction in the nation from implementing any “voting qualification or prerequisite to voting or standard, practice, or procedure...which results in a denial or abridgment of the right of any citizen of the United States to vote on account of race or color.” Today an often challenged voting procedure is the drawing of political districts themselves, on the grounds that a political tool, called gerrymandering, dilutes the effectiveness of the vote.

Gerrymandering is the process of manipulating the boundaries of electoral districts to establish an unfair political advantage for some group of people. This process occurs at all levels of government in the United States of America, and regularly occludes the representative characteristic of the democracy.

Gerrymandering is the abusive version of the more general *redistricting*, which is simply the process by which elected and appointed officials redraw electoral districts. In order to guarantee fair representation, electoral districts ought to be drawn based on data about the population and its distribution in space. There is a consensus that “fairness” requires population balance, hence redistricting is done at least every ten years in the United States, immediately after a census. When redistricting is corrupted, it can become a powerful political tool for diluting or magnifying a demographic group’s voting power in elections. When that demographic group is a race, gerrymandering becomes addressable by the VRA.

Quantifying the effects of voting standards, like the electoral map, on racial groups is notoriously difficult. Both Congress and the Supreme Court have attempted to set aside rules to prove that voting

standards that sound neutral have racially disparate effects.

In *Thornburg v. Gingles* (1986)¹⁸, the Court established that plaintiffs needed to show that:

1. the racial or language minority group is sufficiently numerous and compact to form a majority in a single-member district
2. that group is politically cohesive
3. and the majority votes sufficiently as a bloc to enable it to defeat the minority's preferred candidate

This amounts to two kinds of evidence, now known colloquially as the Gingles Test, needed to successfully prove a violation of the Voting Rights Act:

1. the geographic conditions to draw an effective district
2. racial polarization that blocks the will of minority voters

The racially polarized voting test requires showing that a marginalized racial group is *politically cohesive*, and that the “majority votes sufficiently as a bloc to enable it to usually defeat the minority’s preferred candidate.” Both of these are uniquely hard to show given the secret ballot; that is, all voter choices in an election are anonymous. Thus, it is impossible to access the ground truth about how racial groups in the majority and minority vote. This information is necessary to challenging an electoral map that one believes is gerrymandered.

To fill this gap, statistical methods, most notably King’s Ecological Inference (EI)¹⁹, exist to infer that information from publicly available data.

EI was pioneered in the 1990s and 2000s by Harvard professor Gary King and Columbia professor Andrew Gelman, and it became the standard technique for attempting to analyze racial polarization

in elections in service of mounting VRA litigation. As King himself billed EI, it is “the process of learning about discrete individual-level behavior by analyzing data on groups”¹⁹.

From only the aggregate vote counts and aggregate racial distribution, EI produces inferred candidate preferences for each racial group. This is important because the courts that litigate VRA cases typically do not accept data from non-public sources. Often, this means that the only data that plaintiffs have to bring forward are Census data and election results.

One method of solving the Ecological Inference problem is King’s own Ecological Inference method. The simplest versions of this method use Normal distributions, and the most complex hinge on Binomial-Beta hierarchical models for Bayesian inference. As a point of departure, this paper explores a new way of providing data to the hierarchical model pipeline: by discretizing the whole parameter space of possible vote outcomes. With this discretized space, one can replace the limited Beta class of distributions with a much more flexible description of voter behavior. This new method for solving the Ecological Inference problem, the *discrete voter model* (DVM), was not computationally tractable at the full scale of a U.S. state in the 1990s and early 2000s, but is now, due to advances in algorithm design and computing power.

While gerrymandering is difficult to prove, it is clear that it is one of the most pressing contemporary voting rights problems. The nation’s electorate, and thus the integrity of the democracy, has to constantly contend with confusing and dilutive maps that are drawn by people, parties, and agencies with specific agendas. It is important that the methods for analyzing electoral maps are constantly updated and improved with new knowledge and computing power. Additionally, it is critical that organizers, potential plaintiffs, and members of various affected communities all have the ability to quickly and

intuitively synthesize the work of mathematicians, statisticians, geographers, political scientists, and other key stakeholders towards resolving this longstanding and unavoidably complex problem.

This Court has long understood that it has a special responsibility to remedy violations of constitutional rights resulting from politicians' districting decisions...Part of the Court's role in that system is to defend its foundations. None is more important than free and fair elections.

Supreme Court Justice Elena Kagan

1

Background and Literature Review

I.I RACIAL GERRYMANDERING AND THE COURTS

GERRYMANDERING is the process of manipulating the boundaries of electoral districts to establish an unfair political advantage for some group of people. Racial gerrymandering has been affirmed by the court as a direct violation of constitutional rights, with the Voting Rights Act (VRA) used to

further bolster those rights. A racial gerrymander constitutes a “systemic and infrastructural facet of the practical application of the right to vote,”¹⁶ and plaintiffs often invoke the VRA to force officials to redistrict.

Racially polarized voting, defined as different racial groups voting for different candidates in an election, has been central to redistricting law since the 1982 amendments to section 2 of the Voting Rights Act.¹⁶ Two kinds of evidence, now known colloquially as the Gingles Test, after the 1986 *Thornburg v. Gingles*¹⁷ case, further codified that in order to prove a violation of the Voting Rights Act one needed to show that:

1. the racial or language minority group is sufficiently numerous and compact to form a majority in a single-member district
2. that group is politically cohesive
3. and the majority votes sufficiently as a bloc to enable it to defeat the minority’s preferred candidate

The final two requirements require showing racially polarized voting. “Political cohesion” is often interpreted as racial groups voting for one candidate, and that racial voting information is necessary to determine if racial minority group’s candidates are often defeated by the majority racial group’s candidates.

This means, practically, that in order to win a lawsuit plaintiffs must prove that voters in districts vote along racial lines. While not every instance of racially polarized voting occurs because of a gerrymandered electoral map, proof of racially polarized voting is often the first step in successfully litigating a racial gerrymandering case as a violation of the VRA.

1.1.1 ESTIMATING RACIALLY POLARIZED VOTING

ALTHOUGH racially polarized voting is critically important to racial gerrymandering litigation, it is quite difficult to prove. The *secret ballot*, an essential part of the United States' democracy, ensures that one's vote is anonymous. Hence, any data collection around voting patterns that separates along any axis of identity, is impossible, and all information about those voting patterns must be inferred.

The *Gingles* decision noted that two techniques were “standard in the literature for the analysis of racially polarized voting”⁵: bivariate ecological regression and homogeneous precincts.

Dozens of court cases, from the 1980s to now, have relied on either of the above methods for litigating racial gerrymandering cases¹⁶. However, some appellate courts and judges have been hostile to them. In *Lewis v. Alamance County (1996)*⁴, the court said that the critical assumption of regression “runs counter to common sense.” In *Holder v. Hall (1994)*¹⁷, said that “bivariate regression analysis...does not directly control for...factors [other than race].”

These criticisms were often ignored by peer and other courts because they did not posit new ways of deciding cases without using regression. Chief Justice Roberts chimed into the debate as well, saying “At trials, assumptions and assertions give way to facts. In voting rights cases, that is typically done through regression analyses of past voting records.”³

However, the problem has gotten increasingly more complex. Jim Greiner, Professor of Law at Harvard Law School, stated in his paper *Ecological Inference in Voting Rights Act Disputes: Where Are We Now, and Where Do We Want To Be?*¹⁸

According to some scholars, the patterns themselves have shifted in that a certain percentage of white citizens in some jurisdictions [like Chicago] have become willing to vote for minority-preferred candidates. This shift suggests that the days in which “documenting racially polarized voting is generally...just beating the obvious” may be gone. Racial bloc voting, often a hotly contested issue in an “ordinary” redistricting dispute, may now become even more so, and as a result, using the best methodology is more critical.¹⁶

Greiner’s statement is a good introduction to the future of redistricting. He notes not only that the electorate’s voting patterns are becoming more complex, but also that to ensure that the courts are equipped with the correct tools to tackle the complexity, the methodologies supplied by and to expert witnesses must be continually improved.

The ecological regression and homogeneous precincts methods had several disadvantages. Ecological regression regularly produces impossible estimates of support, like that –400% of Hispanic voters in a district supported a candidate. The analysis of homogeneous precincts depends on the assumption that white people living in an all-white precinct vote similarly to white people living in a more racially-diverse precinct, which has been shown empirically, and can be thought intuitively, to be false.¹⁶ In addition, neither method generalizes well past 2 racial groups and 2 candidates in an election.

The dominant racial bifurcation in both legal and lay discourse is that between Black and white citizens. Thus, much of the litigation of racial gerrymandering and VRA cases, along with the methods used, have focused on these two groups. However, many contemporary cases explore the more complex racial dynamics that exist in our more, and increasingly, diverse electorate.

Take, for example, a recent case that used ecological regression: *League of United Latin American Citizens v. Perry*, 126 S. Ct. 2594 (2006) (LULAC). In this case, Latinx plaintiffs, a mixture of private

citizens and representatives of the Mexican American Legal Defense and Educational Fund, argued that Texas' congressional districts were racially gerrymandered to negatively affect Latinx citizens in Texas. With the help of a dubiously-adjusted ecological regression model (to be used for a larger combination of racial groups), the court decided to throw out District 23 within a new plan proposed in 2003, ordering a remedial redistricting.³ The court also stated that it need not rule that District 25 was a racial gerrymander, given that the remedial changes to District 23 would necessarily affect District 25.

Although not verbally verified by the court, District 25 is widely believed to have been drawn to pack Latinx voters into a district and dilute their voting power, so the redrawing of it and District 23 was a win for that section of the electorate. The change had incredibly practical, and almost immediate, effects. In the year of the remedial redistricting, the affected districts held new primary elections. Henry Bonilla, the Republican representative for the 23rd congressional district, was defeated by Democrat Ciro Rodriguez, a favorite of the Latinx electorate.

Figure 1.1 shows the creation of District 25 after the 2003 redistricting in question. Its shape is reminiscent of popular gerrymanders.

In that decision, the judges referred to a previous one from 1993, where the court stated:

In countless areas of the law weighty legal conclusions frequently rest on methodologies that would make scientists blush. The use of such blunt instruments in examining complex phenomena and corresponding reliance on inference owes not so much to a lack of technical sophistication among judges, although this is often true, but to an awareness that greater certitude frequently may be purchased only at the expense of other values.²

Here, the court explicitly acknowledges the disadvantages of the methods used to infer racially polarized voting. These disadvantages, the criticisms lain on ecological regression and homogeneous

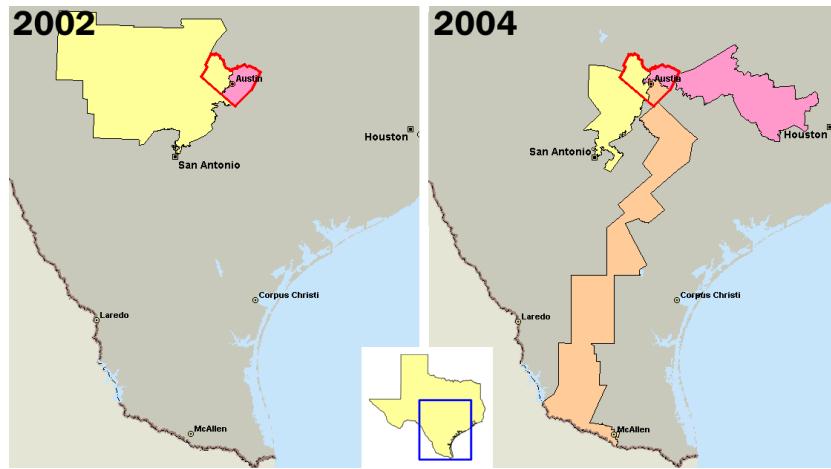


Figure 1.1: The Creation of Texas' 25th Congressional District (orange) in 2003 ¹²

precincts by some courts, and improvements in theory inspired a burst of energy and innovation in the academy in the late 1990s and early 2000s. At the center of this burst is King's EI.

1.2 ECOLOGICAL INFERENCE

ECOLOGICAL INFERENCE (EI) is the process of using aggregate (or, ecological) data to infer discrete individual-level relationships of interests when individual-level data are not available.¹⁹ This is distinct from the *ecological regression* discussed above, which is the statistical method of running regressions on aggregates and interpreting those regressions as predictive relations on the level of individual units.¹⁵

In a world with two categories, groups $\{1, 2, \dots\}$ and groups $\{A, B, \dots\}$, both methods seek to answer the question “what percentage of members in group 1 are also in group A?” This is precisely the answer that plaintiffs need in racial gerrymandering cases, where groups $\{1, 2, \dots\}$ are races and groups $\{A, B, \dots\}$ are candidates.

Table 1.1: A 3×3 Table of Voting By Race, using Counts of Voting Age Population

	Cand. 1	Cand. 2	Σ
Race 1	-	-	Race 1 Pop.
Race 2	-	-	Race 2 Pop.
\sum	Cand. 1 Votes	Cand. 2 Votes	Precinct Pop.

Greiner provides an alternate description of EI that fits well within mathematical models.¹⁶ EI can also be understood as the attempt to predict internal cell values of a set of contingency tables, like Table 1.1, when only the column and row totals (the margins of the table) are observed.

Each precinct has its own table, with an election's vote counts as the column totals and the voting age population, or demographic split of the district, often from the Census, as the row totals. The internal cells, how many people of each racial group that voted for each candidate, are protected by the secret ballot. Those internal cells are exactly the evidence needed in racial gerrymandering litigation, and their values have been inferred by ecological regression and homogeneous precinct analysis in the past.

Theoretically, these tables can be expanded to be as large as necessary: $R \times C$, where R corresponds to the number of rows (demographic groups or races) and C corresponds to the number of columns (candidates). This would be ideal given that the diversification of the electorate has created more politically powerful demographic groups, and that there are frequently more than 2 candidates in elections, especially in local cases (where many infringements of the VRA are suspected to occur).

Professor Gary King developed and proposed his own method to solve this problem in the late 1990s. It has been used in a few court cases, mainly for the advantage that unlike ecological regression,

this method will never produce impossible estimates. It is also understood to be more intuitive than regression, addressing the court's seldom complaint.

1.2.1 KING'S ECOLOGICAL INFERENCE

PROFESSOR KING introduced his method to solve the ecological inference problem in his 1997 book *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*¹⁷. His method, which he expands upon with several papers from 1997 to the late 2000s, uses Bayesian inference and bounds methods to infer the values of the empty cells in Table 1.1.

Using Table 1.1 as a reference, Table 1.2 describes the quantities that King used to describe this problem¹⁸.

Table 1.2: Quantities in King's EI

Quantity	Description
Observed	
X_i	fraction of population i in race 1
T_i	fraction of population i who voted for candidate 1.
N_i	size of population i
T'_i	number of people in population i in race 1
Unobserved	
$b1_i$	fraction of members of race 1 who voted for candidate A
$b2_i$	fraction of members of race 2 who voted for candidate A

King's method then uses the observed quantities, and a few parameters, to find the unobserved quantities and answer the ecological inference question. The parameters used are described in Table 1.3.

Table 1.3: Parameters in King's EI

Parameter	Description
c_1, d_1	hyperparameters for the Beta distribution for b_1
c_2, d_2	hyperparameters for the Beta distribution for b_2 .
θ	the expected fraction of the population of i registered to vote i
λ	the fixed hyperparameter for the distributions of c_1, c_2, d_1, d_2

The model is a hierarchical Bayesian inference model that proceeds as follows:

1. $c_1, c_2, d_1, d_2 \sim \text{Exponential}(\lambda)$
2. $b1_i | c_1, d_1 \sim \text{Beta}(c_1, d_1)$
3. $b2_i | c_2, d_2 \sim \text{Beta}(c_2, d_2)$
4. $T'_i | b1_i, b2_i, X_i \sim \text{Binomial}(N_i, \theta_i)$

The Python implementation of King's EI is given in Listing A.1. Using the observed quantities, King's method produces a model for the unobserved quantities of interest, which can then be used in litigation as described above.

However, in King's original paper, the bulk of his examples are for 2×2 tables, and proposed extensions of that method beyond that bound have been dubious.

In King's second iteration of his method and accompanying software, *EIz*, King implements a “Multiple Imputation Approach” to try to handle 2×3 tables. In essence, this approach collapses the 2×3 table into multiple 2×2 subtables, then applies the original method to each of those subtables.¹⁸ Statisticians and scholars, like Jim Greiner, have noted that this method as applied is statistically invalid, as it does not, and cannot, fully adhere to combinatorial rules that the pioneer of multiple imputation, Professor Donald Rubin of Harvard University, notes in his writings.²⁵

The problems increase if the size of the table increases to 3×3 or higher. Professor Karen Farree of the University of California, San Diego investigated King's EI applied to tables of that size, and found that the method produced biased estimates and relied on inappropriate modeling assumptions.¹⁴

Hence, although King's method has proved quite useful for some cases of ecological inference, with an increasingly diverse electorate, and complex multi-candidate elections, a method that generalizes well to an $R \times C$ table, using the same data available to previous methods, is necessary.

1.3 DISCRETIZATION

KING'S METHOD assumes the quantities of interest, the demographic voting patterns of a district, can be modeled completely with Beta distributions. The hierarchical model allows some freedom to those distributions, generating their parameters from Exponential distributions.

Beta distributions are known to be approximately bimodal if both shape parameters are below 1. However, Beta distributions can never be *fully* bimodal, which is the expectation for racially polarized voting: that two different groups vote quite differently. Figure 1.2 is an example of a very, but not fully, bimodal Beta distribution.

Perhaps more importantly, these distributions can never be multimodal (with more than 2 modes), which is the case that is necessary for multiple racial groups and candidates. This is where discretization becomes quite useful. Figure 1.3 is an example of a true multimodal distribution over two variables.

A distribution like this could be created easily with Neural Networks, but that architecture requires

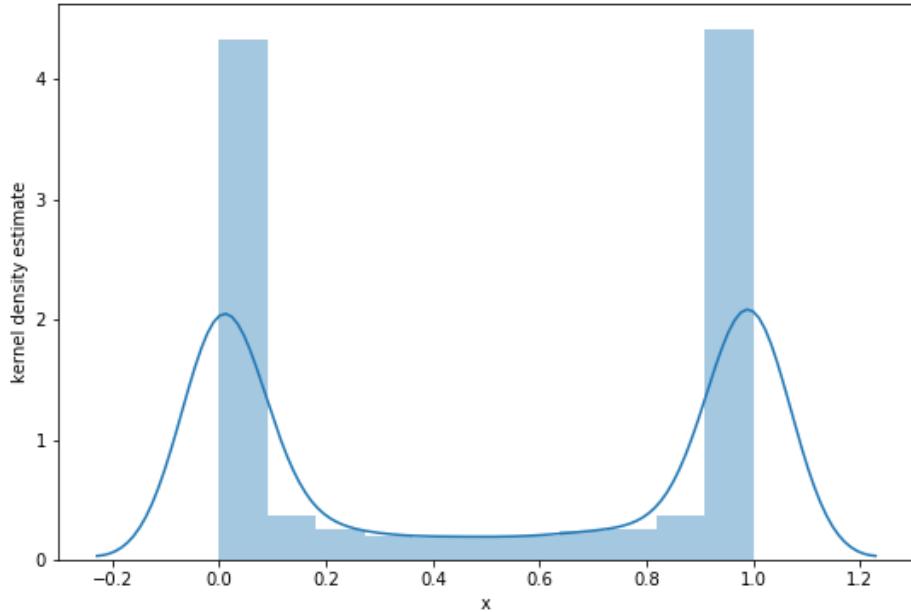


Figure 1.2: Kernel Density Plot of the Beta(0.1, 0.1) Distribution

tens of thousands or millions of data points to be reliable, where a “data point” would probably be an electoral result. The problem of ecological inference, especially in local elections, is simply too small for those methods.

However, by specifying very little about the space (even less than something like a Beta distribution), then allowing the space to take the shape of the available data, more complex, interesting, and possibly accurate distributions can be found. This is called the *Discrete Voter Model (DVM)*.

For a given candidate, each state of this space is a hypercube (the n -dimensional abstraction of a 2-D square/grid or a 3-D cube) of values, where each dimension corresponds to a demographic group of interest. Each cell of the hypercube has a probability, where the sum of the probabilities of all cells

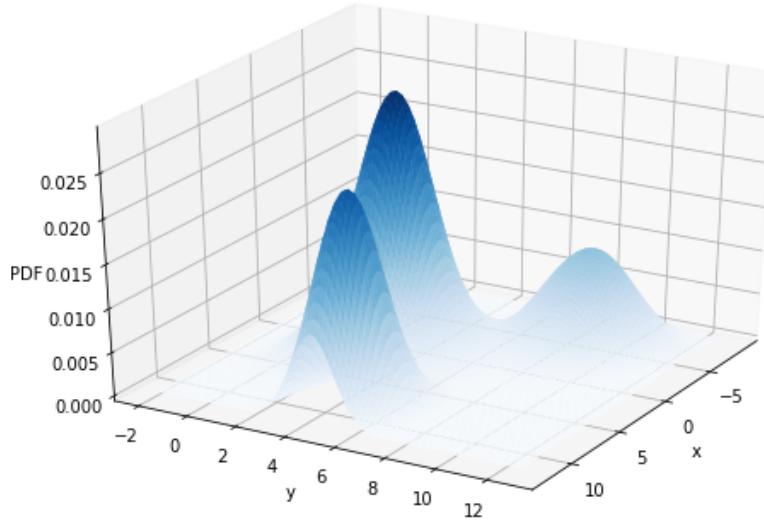


Figure 1.3: A Bivariate Multimodal Distribution

is 1. Each cell also uses its position (the values on each axis) in the full hypercube to represent the demographic voting patterns of a given precinct.

For example, in the 2-dimensional case, for a hypercube with granularity 10 (each axis goes from 0 to 10), let the cell at (2, 5) have an internal probability of 0.4. This corresponds to there being a 40% chance that this candidate's demographic voting pattern in this precinct is:

- 20% of group 1 votes for them
- 50% of group 2 votes for them

Hence, a single hypercube represents the distribution of possible demographic voting patterns in a precinct. Therefore, the distribution of hypercubes represents possibilities for all precincts in a map,

and for an election.

The procedure by which the state space of hypercubes takes the shape of the data is Markov chain Monte Carlo.

1.3.1 MARKOV CHAIN MONTE CARLO

MARKOV CHAIN MONTE CARLO (MCMC) is a class of algorithms that sample from probability distributions that are hard to sample from, typically called *target distributions*.¹¹ With many distributions, like the Normal, Beta, and Exponential mentioned above, obtaining a sample or specifying those distributions is simple and fast. There exist many packages for Python, R, and other programming languages that generate samples instantly with closed form methods.

The target distribution for the Discrete Voter Model, that over the space of hypercubes that could have possibly resulted in the outcome of an election, is not well-specified, and differs not only by election, but also by precinct. Attempting to apply a well-known distribution requires many approximations and assumptions. In lieu of a well-known distribution, the Discrete Voter Model uses discretization and MCMC to find the estimates of the quantities of interest.

All MCMC algorithms are iterative, and the samples converge to the target distribution over time, so the understanding of that target distribution grows with each iteration. All MCMC algorithms construct Markov chains, which means that each sample is generated from, and only depends on, the prior sample in the process, establishing the Markov relationship.²¹ At each step, a candidate for the next sample generated according to some transition rules, determined by a *transition kernel*. That can-

dicate is either accepted or rejected based off of some acceptance probability function. More precisely:

Let the chain be represented by X_0, X_1, \dots, X_t , where X_i is a step in the chain at iteration or time i .

Let the target distribution be $\pi(\cdot)$, which has an unnormalized density π_u .

Let there be some proposal distribution, $Q(\cdot)$, which generates new candidates at every step of the chain. Therefore, as a function $q(x, \cdot)$, it generates Y_{i+1} as a candidate for X_{i+1} , given X_i .

Let there be an acceptance function:

$$\alpha(x, y) = \min \left[1, \frac{\pi_u(y) \cdot q(y, x)}{\pi_u(x) \cdot q(x, y)} \right]$$

The candidate state Y_{i+1} is accepted with the probability $\alpha(X_i, Y_{i+1})$. If the state is accepted, $X_{i+1} \leftarrow Y_{i+1}$. If not, $X_{i+1} \leftarrow X_i$.

The differences in transition kernels, $Q(x, \cdot)$, and acceptance probability functions, $\alpha(x, y)$, are they key delineators of different types of MCMC algorithms. The Discrete Voter Model uses two types of transition kernels: *Random Walk Metropolis* and *Hamiltonian Monte Carlo*.

1.3.2 RANDOM WALK METROPOLIS

ONE MARKOV CHAIN MONTE CARLO METHOD which uses the Random Walk Metropolis transition kernel is the Metropolis-Hastings algorithm. Metropolis-Hastings is one of the simplest MCMC methods, but is still quite reliable and widely used for many applications.

This transition kernel requires that $q(x, y) = q(y - x)$. This is one way to ensure that the chain

is reversible and detail balanced, necessary conditions for the chain to converge to the target distribution.²⁴. Examples of such distributions are $Q(x, \cdot) = N(x, \sigma^2)$ and $Q(x, \cdot) = Uniform(x - 1, x + 1)$.

The Discrete Voter Model's proposal randomly selects two cells in the probability hypercube, and adds a small value, ε , to one and subtracts it from the other. This is a reversible chain – one could get theoretically get back to the probability hypercube that was changed later in the chain – which is essential for convergence.¹¹

1.3.3 HAMILTONIAN MONTE CARLO

HAMILTONIAN MONTE CARLO (HMC) differs from Random Walk Metropolis in that its proposal distribution adapts to the current state and its position in the parameter space.¹⁰

HMC uses the gradient of the target distribution's density, and adapts the proposal distribution toward that direction. This helps combat some of the inefficiencies of Random Walk Metropolis, like its tendency to explore parts of the state space where there is very little probability density. Hence, fewer samples have to be created to approximate the target distribution.

The “Hamiltonian” in “Hamiltonian Monte Carlo” refers to the inspiration of this kernel: Hamiltonian mechanics, a reformation of Newtonian mechanics that reconsiders the basis of time evolution of a system.¹³ This kernel generates some *potential function/energy* from the target density, then uses the curvature of that function, some random seed, and the current state to determine the candidate for the next state.²³

The primary benefit of HMC to the Discrete Voter Model is that it can theoretically produce a

better sample of the target distribution with fewer steps in the algorithm. In addition, HMC has been shown to work well on state spaces with many parameters, like the potentially large probability hypercubes in this model.²³

1.3.4 ROBUSTNESS

THE MARKOV CHAIN CENTRAL LIMIT THEOREM guarantees convergence and accurate sampling with the correct specifications.²⁴ Furthermore, for both the kernels above, chains are robust to their assumptions – they often converge to reliable samples in enough time, even when assumptions (like a perfectly ergodic chain) are broken.

The main disadvantages of using Markov chains for such complex state spaces are that it is difficult to assess convergence, and the methods themselves are computationally expensive. However, given the increasing availability of computing power, including the prevalence of GPU-based computing and the maturity of programming packages like Stan, PyMC3, and TensorFlow Probability, these methods can be more widely applied.

We should forget about small efficiencies, say about 97% of the time: premature optimization is the root of all evil. Yet we should not pass up our opportunities in that critical 3%.

Dr. Donald Knuth

2

Methods

The Discrete Voter Model is implemented in Python 3, and the accompanying code is found in the repository noted in the [Appendix](#). A set of 5 modules supports the Discrete Voter Model, with each implementing an aspect of the statistical framework. Those modules are:

1. `phc`: creates probabilistic hypercubes
2. `expec_votes`: finds the expectation of votes for a candidate

3. `prob_votes`: finds the probability of an electoral outcome
4. `elect`: aggregates election and electorate data
5. `dvm`: executes the Markov chain on a state space of hypercubes

The modules work together to run the Discrete Voter Model on an `Election` object to generate a distribution of probabilistic hypercubes for the election. Each probabilistic hypercube encodes a probability distribution within a potential precinct of the district, and is used to extract the likely demographic voting probabilities.

These demographic voting probabilities are the goal of ecological inference, and provide information to answer whether voting in an election is polarized by race.

2.1 PHC

THE DISCRETE VOTER MODEL introduces the probabilistic hypercube (PHC) to represent the distribution of possible demographic voting probabilities in a precinct, for a candidate.

A probabilistic hypercube (PHC) is defined as a rank n tensor (matrix, multidimensional array, or holor) in \mathbb{R}^n space with the following characteristics:

1. the component values of the cells sum to 1
2. each dimension corresponds to a demographic group in an electorate
3. each cell's position, or index, represents the demographic voting probabilities of a precinct
4. each cell's component value represents the inferred probability of that cell representing the true

demographic voting probability of a precinct

The function `make_phc` initializes a PHC with either random or uniform cell components. The rank of the PHC corresponds to the number of demographic groups being analyzed by the model. The size of each dimension is the *granularity* of the discretization, and thus the model.

For example, a PHC with granularity 100 meant to represent 3 demographic groups will have the shape $(100, 100, 100)$, and can be thought of as a $100 \times 100 \times 100$ matrix.

The demographic voting probabilities (DVP), the collection of b_i from King's model specification, are derived for a cell directly from that cell's position within the PHC. The higher a cell is along a dimension of the tensor, the higher the probability that members of that dimension's corresponding demographic group will vote for a candidate.

For example, in a precinct with 2 demographic groups, and in a PHC with granularity 10, a cell at the position $10, 0$ represents $\frac{10}{10} = 100\%$ of demographic group 1 voting for the candidate, and $\frac{0}{10} = 0\%$ of demographic group 2 voting for the candidate.

In general, in some PHC with granularity d representing n demographic groups, for some cell at index $\{i_0, i_1, i_2, \dots, i_n\} = \{i_j\}_{j=0}^n$:

$$\text{DVP} = \left\{ \frac{i_j}{d} \right\}_{j=0}^n$$

When given a precinct, candidate, and some PHC, one can determine the demographic voting probability of the precinct for the candidate by sampling a cell, or set of cells, from the PHC randomly, using the cell's component values as probabilities. The higher the component value of a cell, the more

likely it is to be chosen to be the inferred demographic voting probability of the given precinct and candidate.

The granularity of the discretization is directly related to how precise the demographic voting probabilities are. In general, the higher the granularity of a PHC, the more precise its predictions may be. However, higher granularity PHCs have many more cells, and thus parameters, making sampling time/computing intensive.

Figure 2.1 is a rendering of a rank 3, granularity 10 PHC, with the dimensions corresponding to the Latinx, white, and Black members of an electorate. Figure 2.2 is a similar rendering, but of a rank 2, granularity 10 PHC, with the dimensions corresponding to the white and Black members of an electorate. In each, the cells are colored according to their component values, or probabilities – the more yellow and opaque a cell, the higher the probability is in that cell.

The rank 2 case can also be visualized as a 3-dimensional plot, where the third dimension is the probability associated with the cells of the PHC. Figure 2.3 illustrates this.

Higher rank PHCs are difficult to visualize in similar ways, but the representational and intrinsic characteristics of the object remain the same. This is critical to the Discrete Voter Model's ability to generalize to arbitrarily diverse electorates, with arbitrarily precise demographic voting probabilities.

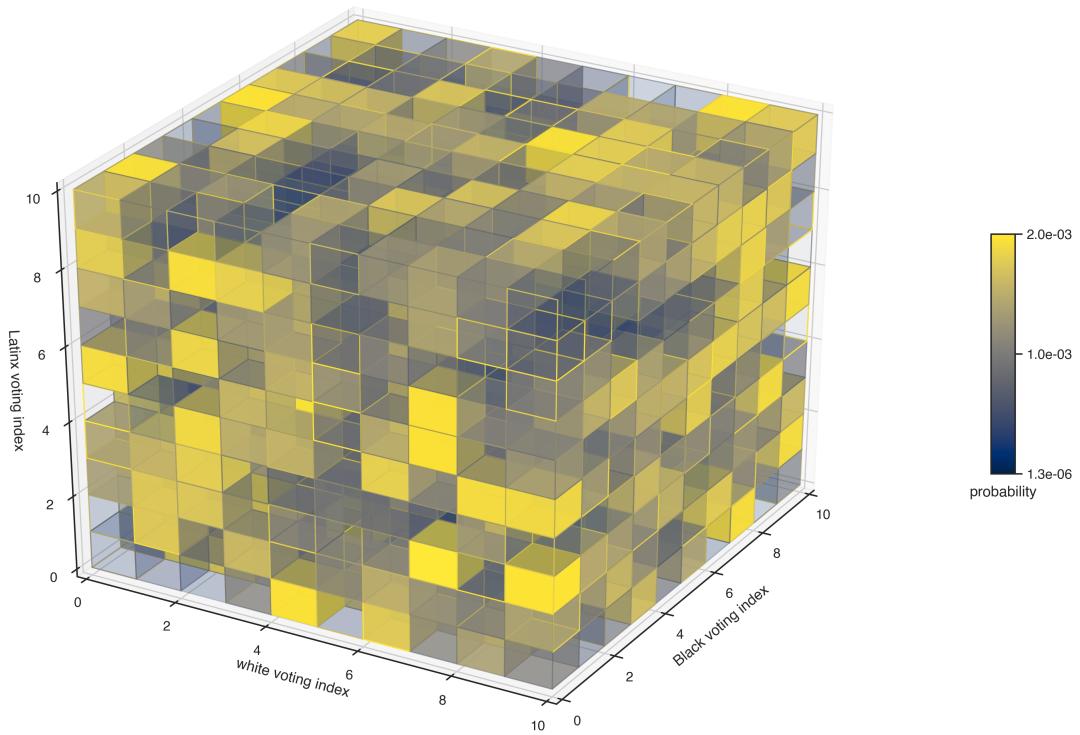


Figure 2.1: Example of a Rank 3, Granularity 10 Probabilistic Hypercube

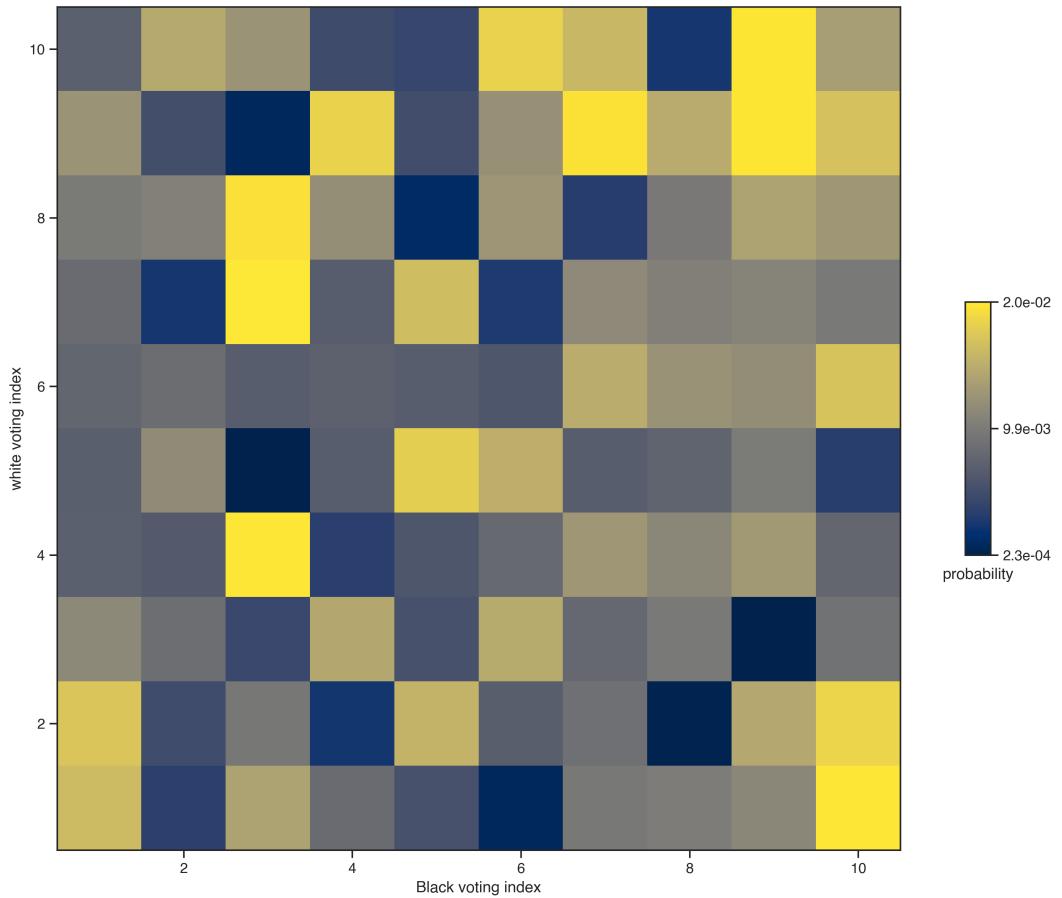


Figure 2.2: Example of a Rank 2, Granularity 10 Probabilistic Hypercube

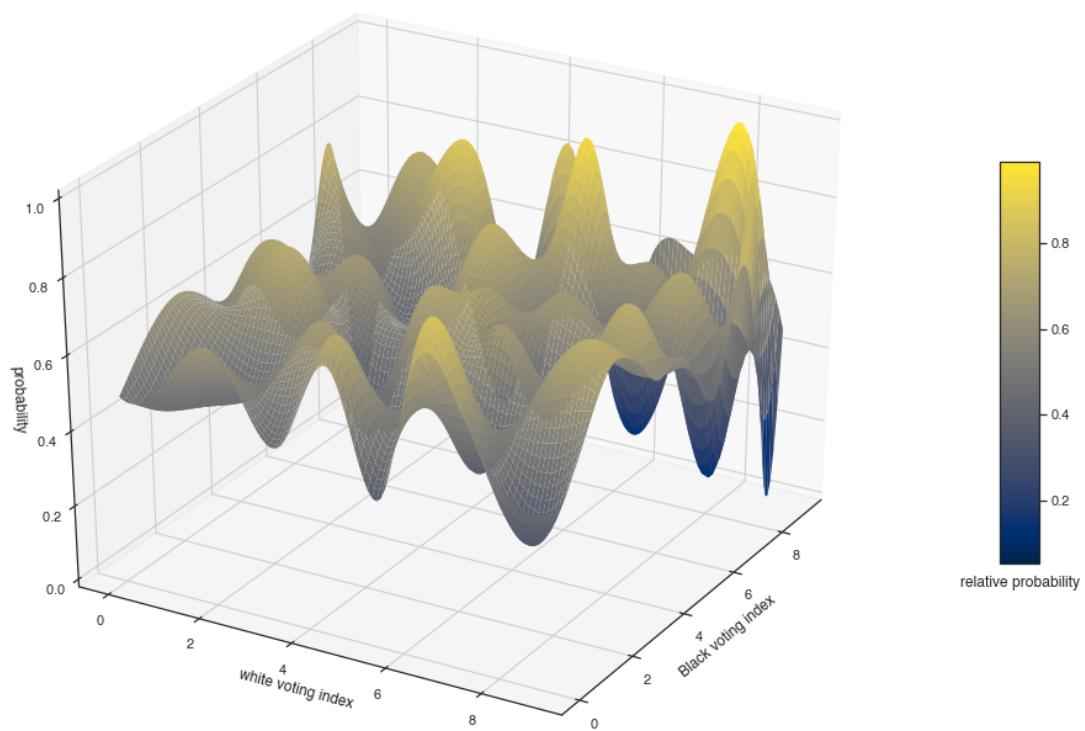


Figure 2.3: The Probability Distribution within of a Rank 2, Granularity 10 Probabilistic Hypercube

2.2 EXPEC_VOTES

TO BE ABLE TO USE PROBABILITY HYPERCUBES as states in a Markov chain Monte Carlo method, the Discrete Voter Model employs two methods of determining how “good” each PHC is. This determination, or scoring, is done with the modules `expec_votes` and `prob_votes`.

`expec_votes` calculates the expectation for the vote outcome from a PHC, and compares that outcome to the observed electoral outcome. The closer they are, the higher the probability that this PHC represents a probable set of demographic voting probabilities for the given precinct.

Let there be some PHC with granularity d representing n demographic groups, for some cell at index $\{i_0, i_1, i_2, \dots, i_n\} = \{i_j\}_{j=0}^n$:

Let $i = 0, 1, \dots, m$ be the cells in the hypercube, where $m = d^n$.

Let p_i be the probability of being in cell i .

Let $j = 0, 1, \dots, n$ be the demographic groups.

Let q_j be the probability of demographic group j to vote for the given candidate. $q_{j=0}^n$ = the demographic voting probabilities from Section 2.1.

Let s_j be the population of demographic group j in the precinct.

Let v be the observed number of votes cast for the candidate in the precinct.

The expectation of a PHC is thus given by Equation 2.1.

$$\mathbb{E}(\text{PHC}) = \sum_{i=0}^m p_i \left(\sum_{j=0}^n q_j \cdot s_j \right) \quad (2.1)$$

In addition to calculating the expectation of a PHC, the `expec_votes` module compares that expectation to the observed electoral vote outcome in a precinct. It calculates the $L1$ -norm of the difference between the two, then uses the function in Equation 2.2 to convert it to a representation of the log-probability that the PHC has the same expectation as the vote outcome.

Let x be the $L1$ -norm of the difference between the expectation of a PHC and the electoral vote outcome, v :

$$x = \|\mathbb{E}(\text{PHC}) - v\|_1$$

$$f(x) = \log \left(\frac{1}{x+1} \right) \quad (2.2)$$

The computation above is repeated for every precinct in an election. Let X be the vector $\{x_i\}_{i=0}^p$ where p is the number of precincts. f is then applied to X , and the average of that quantity, $\overline{f(X)}$, is the score of the PHC.

The more negative the value of $\overline{f(X)}$, the less likely the PHC is to be able to generate the outcome of the election and be accepted in the Markov chain.

2.3 PROB_VOTES

THE DISCRETE VOTER MODEL also uses the `prob_votes` module to score a probability hypercube. `prob_votes` directly calculates the probability that a given hypercube produced the observed election outcome.

With the same notation as `expec_votes` in Section 2.2, additionally:

Let a_j be the number of people in demographic group j that voted for the given candidate.

Equation 2.3 is the probability that a given hypercube produced a given election outcome.

$$\mathbb{P}(\text{PHC} \rightarrow v) = \sum_{\forall a_j \text{ s.t. } \sum_{j \in [0, n]} a_j = v} \left(\prod_{j=0}^n q_j^{a_j} \cdot (1 - q_j)^{s_j - a_j} \right) \cdot \frac{v!}{\prod_{j=0}^n a_j!} \quad (2.3)$$

Equation 2.3 uses the Binomial distribution to calculate the probability of members of different demographic groups voting together in a way to produce the desired outcome. This is extensible to any number of demographic groups and can be repeated for any number of candidates.

A crucial step in this process is generating sets of $\{a_j\}_{j=0}^n$ such that:

$$\sum_{\forall a_j \text{ s.t. } \sum_{j \in [0, n]} a_j = v}$$

These sets are all possible partitions of the observed electoral outcome into the demographic groups.

For example, if 10 people voted for a candidate, and there are three demographic groups, 3 possible partitions for the people who voted for the candidate are shown in Table 2.1.

Table 2.1: Partitions of 10 votes among 3 demographic groups

Partition	Group 1 Votes	Group 2 Votes	Group 3 Votes
1	4	3	3
2	2	0	8
3	2	4	4

Integer partitioning can be computationally expensive. The time complexity of most algorithms for it is $O(v(\log v))$, where v is the vote outcome of an election. However, the Discrete Voter Model limits the size of partitions to be the number of demographic groups in the electorate, n . While DVM generalizes well to a large number of demographic groups, n is unlikely to get large enough to make this integer partitioning step a time bottleneck. With that said, most algorithms for this process have space complexity $O(n^2)$. While limiting the size of partitions does reduce the space necessary, to further mitigate that use of space, DVM creates the set of integer partitions as a Python generator. Hence, values are treated like a stream, where computation is delayed until a partition is necessary, and the set of partitions is treated like a regular iterator. Finally, since integer partitions are used several times in the computation, and only depend on the vote outcome and the shape of the PHC, not the values in the PHC, the partitions can be generated once, and reused at every step in the MCMC method.

The full code for this partitioning process can be found in the appendix, either in the full repository or as Listing A.2.

The partitions are used to calculate the probability, as shown in Equation 2.3. That probability is that of some candidate's and precinct's hypercube producing the observed electoral outcome in a precinct. `prob_votes` uses this to calculate the log-probability.

Like with `expec_votes`, the computation is repeated for every precinct in an election, then aver-

aged over all precincts. The more negative that value, the less likely the PHC generated the outcome of the election and is accepted in the Markov chain.

This scoring method is more intuitive and perhaps more precise, but it is much more computationally expensive. The calculations of large products, exponents, and especially factorials makes this much slower than scoring by expectation with `expec_votes`.

2.4 ELECT

THE DISCRETE VOTER MODEL operates on data from elections, and the `elect` module prepares and packages the data for use.

This module implements the `Election` class, a data structure and accompanying methods that represents an election in a district. It has the following attributes:

1. `candidates`: the candidates
2. `dpp`: demographic distribution, per precinct
3. `dvp`: the demographic voting probabilities, per precinct
4. `mock`: whether the election is a mock election
5. `name`: the name of the election
6. `num_demo_groups`: the number of demographic groups in the district
7. `outcome`: the vote outcome, by demographic group, of the election, per precinct
8. `precincts`: the precinct IDs

9. `vote_totals`: the vote totals of each candidate, per precinct

10. `winner`: the winner of the election

An `Election` object for a mock election is initialized by providing the candidates, demographic distribution per precinct, and demographic voting probabilities (DVP) per precinct. Upon initialization, an election is simulated based on how each demographic votes, and other attributes, like the outcome, vote totals, and winner, are calculated.

That DVP is typically the desired result of the Discrete Voter Model, and by providing it to mock elections that then use it to generate election data, one can run the model on the object and attempt to recover the “truth” that generated the results of the election. This is covered in depth in Section 2.6.

An `Election` object for a real election is initialized by providing the candidates, demographic distribution per precinct, and vote totals per precinct. Upon initialization, the election is “decided” by calculating which of the candidates won.

The `elect` module also provides a function, `create_elections`, that takes demographic data per precinct and voting data per precinct, both as the ubiquitous `Pandas DataFrames`, to create an `Election` object. This allows DVM to be deployed quickly and easily to real election data. Section 2.6 employs it for the case studies of North Carolina and Chicago.

The Discrete Voter Model uses an `Election` object’s candidates, demographic distribution per precinct, and vote totals to create probabilistic hypercubes and run the Markov chain Monte Carlo method to find a distribution of PHCs that could have resulted in that `Election`’s outcomes. For mock elections, the accuracy of DVM can be compared to the original DVP that generated the `Election` object.

2.5 DVM

THE FINAL MODULE, and the one that runs the entire process, is `dvm`. It implements the Markov chain Monte Carlo method with Hamiltonian Monte Carlo (HMC) and Random Walk Metropolis (RWM) kernels, with both probability-based and expectation-based state scoring.

The Discrete Voter Model's MCMC method used probabilistic programming Python package [TensorFlow Probability](#) (TFP) as a substrate. TFP is a relatively new probabilistic programming package maintained by Google and the TensorFlow team. It is built on their popular open source machine learning platform, TensorFlow, and offers many advantages over a pure Python implementation of MCMC.²⁰ The top three advantages to the Discrete Voter Model are:

1. Nearly all of the operations are vectorized, allowing for much more efficient computation on high-dimensional objects like probabilistic hypercubes.
2. It automatically scales to different platforms and architectures. Without any difference in the code, this allows the Discrete Voter Model to be able to be run on CPUs (central processing units) found in most computers, GPUs (graphics processing units) found in more powerful computers and sometimes optimized for machine learning, and TPUs (tensor processing units) developed by Google as AI accelerator application-specific integrated circuits (ASIC). As one increases the granularity of their PHC or the complexity of an election, they will be able to benefit from the increased computing power of modern machines.
3. TensorFlow, and thus TensorFlow Probability, has lazy evaluation built in, through the `Graph`

structure. TensorFlow **Graphs** employ the dataflow programming paradigm to increase the efficiency of parallel computations.²⁶ This allows for:

- (a) easy identification and execution of operations that can be computed in parallel.
- (b) distributed computing to take advantage of several machines at once.
- (c) saving the model as a language-agnostic **Graph**. The **Graph** can then be executed in any of the many TensorFlow-compatible languages. For example, a graph generated in Python can be executed in C++.
- (d) visualizing the **Graph** with built-in tools.

Specifically, TFP is used to create the transition kernels and sample from the Markov chain.

The final subroutine, `mcmc`, runs the Markov chain Monte Carlo method with the Metropolis-Hastings algorithm, employing either `expec_votes` or `prob_votes` to score candidates.

The pseudocode for the MCMC is as follows:

1. Choose a kernel: either Random Walk Metropolis or Hamiltonian Monte Carlo
2. Choose a scoring function, g : either `expec_votes` or `prob_votes`
3. Initialize some probabilistic with `phc`, X_i where $X_{i=0}^t$ is the full sample. Score X_i it with the kernel of choice to get s_i
4. Iterate t times. At each iteration, i :
 - (a) Generate a candidate hypercube, Y_{i+1} , with the kernel
 - (b) Calculate the score for that candidate, $s_{i+1} = g(Y_{i+1})$
 - (c) Accept X_i with probability $\alpha = \frac{s_{i+1}}{s_i}$
 - i. If accepted, $X_{i+1} \leftarrow Y_{i+1}$

- ii. If rejected, $X_{i+1} \leftarrow X_i$
- 5. Burn a predetermined fraction of the observations at the beginning of the sample. Removing that early “burn-in” period has been shown by some studies to be effective in creating a better sample.²¹
- 6. Output the sample as the distribution of probability hypercubes.

The HMC kernel, `tfp.mcmc.HamiltonianMonteCarlo`, is initialized with an adaptive step size for the first 60% of the burn-in period. The adaptive step size is controlled by

`tfp.mcmc.SimpleStepSizeAdaptation`, and is based on research by Andrieu and Thoms that shows that when the first few steps of an HMC-directed chain are allowed to adapt the step size towards the direction of the log probability of the state, the succeeding steps are more fruitful.⁹.

The RWM kernel, `tfp.mcmc.RandomWalkMetropolis`, is initialized with a proposal function that chooses two cells of the probability hypercube at random, and adds a small value, ε , to one while subtracting it from the other. That proposal function is reversible and allows for good mixing, while also being simple enough to be performant.

Both kernels trace the log-probability and log-acceptance rate of the chain during sampling.

`dvm` implements the lower level functions for running the MCMC, `rwm` and `hmc`. It also implements the higher level function `dvm_election` that applies the Discrete Voter Model to an `Election` object.

Finally, the module implements functions, `chain_mle` and `mean_phc`, for finding the Maximum Likelihood Estimate and the sample mean of distributions/samples produced by the model. These are quite useful when determining the central tendencies of the sample, and evaluating it easily.

2.6 EVALUATION

THERE ARE TWO AUXILIARY MODULES, `dvm_plot` and `dvm_eval`, that provide visualizations and analysis of the Discrete Voter Model’s results and behavior.

`dvm_plot` provides a function to generate trace plots of the chains’ results, as well as a bevy of functions to visualize the distribution of PHCs and the PHCs themselves. Those functions generated Figures 2.1, 2.2, and 2.3.

All of the methods for inferring racially polarized voting noted in Chapter 1, including ecological regression, homogeneous precincts, and King’s EI, are necessarily imperfect – as is the Discrete Voter Model. The extent of that imperfection, however, can be evaluated and compared.

`dvm_eval` is crucial to evaluating the Discrete Voter Model’s performance generally, as well as how PHC granularity, the number of MCMC iterations, and kernel and score function choices interact with differences in election sizes and types to produce different results. Those results can be compared to the mock `Election` object’s `dvp` attribute which generated the election’s data. The `dvm_evaluator` function does this comparison by running DVM and calculating the mean squared error (MSE), a common loss metric, between DVM’s result and the `Election`’s DVP. The lower the MSE, the more accurate the model’s predictions. `dvm_evaluator` also records the time taken to run the Discrete Voter Model.

In addition, for the 2×2 case, `dvm_eval` has the function `kei_evaluator` that can similarly evaluate the performance of the canonical King’s Ecological Inference method on mock elections. Those

results can then be compared to the Discrete Voter Model's.

A set of experiments were created and run to determine the viability of the Discrete Voter Model.

Those experiments' specifications and results are in [Results](#).

The oppressors maintain their position and evade responsibility for their own actions. There is a constant drain of energy which might be better used in redefining ourselves and devising realistic scenarios for altering the present and constructing the future.

Audre Lorde

3

Results

SEVERAL EXPERIMENTS were designed to evaluate the performance of the Discrete Voter Model. Two case studies, on elections in Chicago, were chosen to demonstrate DVM's performance on real data. The case studies are discussed in the [Discussion](#), and the experiments' specification and results are below.

3.1 EXPERIMENT SPECIFICATION

THE DISCRETE VOTER MODEL was evaluated on the 2×2 case (i.e. 2 candidates and 2 demographic groups) and the 3×2 case that was untenable for previous methods of determining racially polarized voting.

The Discrete Voter Model was run on mock `Election` objects created with the specifications in Subsections 3.1.1 and 3.1.2, using `dvm_evaluator`. All other variables are in Table 3.1.

Table 3.1: Non-Election Variables for DVM Experiments

Variable	option 1	option 2
number of MCMC steps, t	1000	10000
granularity of the PHCs, g	10	100
scoring function	<code>expec_votes</code>	<code>prob_votes</code>
kernel	Random Walk Metropolis	Hamiltonian Monte Carlo

Not all combinations of variables proved feasible for the scope of these experiments. Some, like using the Hamiltonian Monte Carlo kernel with the `prob_votes` scoring function for 10000 steps on a set of granularity 100 PHCs, were infeasible. The combinations that were run are presented in Section 3.2.

3.1.1 2×2

FOR THIS CASE, 6 mock elections were created for single-precinct districts of size 100, with 2 demographic groups, g_1 and g_2 , and two candidates, a and b .

Each of the 6 has one of the 3 demographic distributions of the district in Table 3.2. Each of these distributions is labeled $\{d_i\}_{i=1}^3$ for reference in Section 3.2.

Each of the 6 has one of the 2 demographic voting probabilities for candidates a and b in Table 3.3.

Each of these voting patterns is labeled $\{v_i\}_{i=1}^2$ for reference in Section 3.2.

The demographic voting probabilities and demographic distributions were combined to create the 6 unique elections.

Table 3.2: Demographic Distributions Tested in the 2×2 Case

Label	$g_1 (\%)$	$g_2 (\%)$
d_1	50	50
d_2	25	75
d_3	10	90

Table 3.3: Demographic Voting Probabilities Tested in the 2×2 Case

Label	Group	$a (\%)$	$b (\%)$
v_1	g_1	50	50
	g_2	30	70
v_2	g_1	25	75
	g_2	20	80

These elections were also used to evaluate King's Ecological Inference method, using the `kei_evaluator` function.

3.1.2 3×2

FOR THIS CASE, 6 mock elections were created for single-precinct districts of size 100, with 3 demographic groups, g_1 , g_2 and g_3 , and two candidates, a and b .

Each of the 6 has one of the 3 demographic distributions of the district in Table 3.2. Each of these distributions is labeled $\{d_i\}_{i=4}^6$ for reference in Section 3.2.

Each of the 6 has one of the 2 demographic voting probabilities for candidates a and b in Table 3.3.

Each of these voting patterns is labeled $\{v_i\}_{i=3}^4$ for reference in Section 3.2.

The demographic voting probabilities and demographic distributions were combined to create the 6 unique elections.

Table 3.4: Demographic Distributions Tested in the 3×2 Case

Label	g_1 (%)	g_2 (%)	g_3 (%)
d_4	50	50	0
d_5	25	25	50
d_6	33	33	34

Table 3.5: Demographic Voting Probabilities Tested in the 3×2 Case

Label	Group	a (%)	b (%)
v_3	g_1	20	80
	g_2	80	20
	g_3	40	60
v_4	g_1	30	70
	g_2	40	60
	g_3	50	50

3.2 EXPERIMENT RESULTS

THE `dvm_evaluator` and `kei_evaluator` functions report two metrics: accuracy and runtime.

The “accuracy” of the model on a mock election is the average mean squared error (MSE) between the

model's outputted demographic voting probabilities (DVP) and the DVP that generated the election data, stored in the `Election` objects. The MSE is averaged over all 5 trials of the experiment. The lower the MSE, the more accurate the model. The “runtime” of the model is how long, in seconds, the model took to generate its output. The lower the runtime, the faster the model.

Tables 3.6 to 3.11 summarize the results of running the Discrete Voter Model and King’s Ecological Inference method on all of the experiments, for a total of 66 experiments replicated 5 times each. Each sample had a burn-in fraction of 0.3. HMC configurations had a step size adaptation fraction of 0.6.

A select number of figures, comparing the performance of different experiment configurations on the same election in the 2×2 and 3×2 cases, are shown in Subsections 3.2.6 and 3.2.7. Other such figures for all elections can be found in the [Appendix](#).

In each table, g is the granularity of the probability hypercube, and t is the number of steps.

The experiments were conducted in Python 3.7.4 on a 2018 MacBook Pro with a 2.3 GHz Quad-Core Intel Core i5 processor, 16 GB of 2133 MHz LPDDR3 RAM, and a Intel Iris Plus Graphics 655 1536 MB graphics card. All numbers are given with 4 decimal places of precision.

3.2.1 RANDOM WALK METROPOLIS IN THE 2×2 CASE

Table 3.6: Experiment Results for the RWM Kernel, Scoring by Expectation, for 2×2 Elections

Label	Time (s)	MLE PHC MSE	Mean PHC MSE	g	t
$d_1 \times v_1$	17.6172	0.1158	0.1825	10	1000
$d_1 \times v_2$	17.6489	0.0362	0.0463	10	1000
$d_2 \times v_1$	16.7197	0.1258	0.2092	10	1000
$d_2 \times v_2$	16.5030	0.1062	0.1021	10	1000
$d_3 \times v_1$	16.6309	0.0392	0.0434	10	1000
$d_3 \times v_2$	17.1568	0.5262	0.5361	10	1000
$d_1 \times v_1$	733.3777	0.0739	0.0831	100	1000
$d_1 \times v_2$	725.9847	0.0111	0.0126	100	1000
$d_2 \times v_1$	727.2388	0.1298	0.2143	100	1000
$d_2 \times v_2$	722.3091	0.4098	0.4210	100	1000
$d_3 \times v_1$	729.2840	0.0633	0.0323	100	1000
$d_3 \times v_2$	721.8541	0.0018	0.0121	100	1000
$d_1 \times v_1$	153.4124	0.0825	0.0258	10	10000
$d_1 \times v_2$	151.0056	0.0062	0.0032	10	10000
$d_2 \times v_1$	150.8916	0.1225	0.0692	10	10000
$d_2 \times v_2$	150.7159	0.0813	0.0912	10	10000
$d_3 \times v_1$	150.4812	0.0458	0.0471	10	10000
$d_3 \times v_2$	151.1887	0.0362	0.0217	10	10000

Table 3.7: Experiment Results for the RWM Kernel, Scoring by Probability, for 2×2 Elections

Label	Time (s)	MLE PHC MSE	Mean PHC MSE	g	t
$d_1 \times v_1$	278.7961	0.2125	0.2327	10	1000
$d_1 \times v_2$	121.7805	0.1462	0.3412	10	1000
$d_2 \times v_1$	161.3856	0.0625	0.0594	10	1000
$d_2 \times v_2$	143.3614	0.0162	0.0211	10	1000
$d_3 \times v_1$	85.4510	0.0625	0.0611	10	1000
$d_3 \times v_2$	73.7169	0.2413	0.2209	10	1000
$d_1 \times v_1$	26206.2664	0.1168	0.1004	100	1000
$d_1 \times v_2$	8688.8182	0.0876	0.0914	100	1000
$d_2 \times v_1$	14839.9089	0.0522	0.0514	100	1000
$d_2 \times v_2$	12406.1474	0.0352	0.0446	100	1000
$d_3 \times v_1$	6496.2049	0.0781	0.0698	100	1000
$d_3 \times v_2$	6171.4175	0.1921	0.1810	100	1000
$d_1 \times v_1$	2698.0366	0.2525	0.2731	10	10000
$d_1 \times v_2$	1190.9839	0.0062	0.0101	10	10000
$d_2 \times v_1$	1496.4504	0.1325	0.1052	10	10000
$d_2 \times v_2$	1418.3957	0.0062	0.0172	10	10000
$d_3 \times v_1$	874.6798	0.0925	0.0831	10	10000
$d_3 \times v_2$	820.0640	0.1712	0.1322	10	10000

3.2.2 HAMILTONIAN MONTE CARLO IN THE 2×2 CASE

Table 3.8: Experiment Results for the HMC Kernel for 2×2 Elections

Label	Time (s)	MLE PHC MSE	Mean PHC MSE	g	t	scorer
$d_1 \times v_1$	2363.7653	0.1225	0.1119	10	1000	prob
$d_1 \times v_2$	971.0193	0.2762	0.2339	10	1000	prob
$d_2 \times v_1$	1253.7283	0.0658	0.0392	10	1000	prob
$d_2 \times v_2$	1113.9393	0.1979	0.2879	10	1000	prob
$d_3 \times v_1$	710.3838	0.0625	0.0851	10	1000	prob
$d_3 \times v_2$	641.3496	0.0212	0.0311	10	1000	prob
$d_1 \times v_1$	66.8853	0.0725	0.0645	10	1000	expec
$d_1 \times v_2$	66.8914	0.3262	0.2977	10	1000	expec
$d_2 \times v_1$	66.5275	0.1505	0.1025	10	1000	expec
$d_2 \times v_2$	66.4892	0.0512	0.0331	10	1000	expec
$d_3 \times v_1$	66.7315	0.0125	0.0251	10	1000	expec
$d_3 \times v_2$	68.7194	0.0212	0.0327	10	1000	expec

3.2.3 KING'S ECOLOGICAL INFERENCE METHOD IN THE 2×2 CASE

Table 3.9: Experiment Results for King's Ecological Inference Method on 2×2 Elections

Label	Time (s)	MSE	t
$d_1 \times v_1$	17.5591	0.0246	1000
$d_1 \times v_2$	25.9120	0.0014	1000
$d_2 \times v_1$	14.0912	0.0013	1000
$d_2 \times v_2$	23.2465	0.0523	1000
$d_3 \times v_1$	13.7835	0.0014	1000
$d_3 \times v_2$	13.7010	0.0332	1000

3.2.4 RANDOM WALK METROPOLIS IN THE 3×2 CASE

Table 3.10: Experiment Results for the RWM Kernel, Scoring by Expectation, for 3×2 Elections

Label	Time (s)	MLE PHC MSE	Mean PHC MSE	g	t
$d_4 \times v_3$	81.2624	0.1536	0.1269	10	1000
$d_4 \times v_4$	77.8610	0.0647	0.0869	10	1000
$d_5 \times v_3$	75.6234	0.1225	0.1669	10	1000
$d_5 \times v_4$	76.2014	0.1447	0.0425	10	1000
$d_6 \times v_3$	75.4726	0.0714	0.0958	10	1000
$d_6 \times v_4$	75.7152	0.0781	0.0936	10	1000

3.2.5 HAMILTONIAN MONTE CARLO IN THE 3×2 CASE

Table 3.11: Experiment Results for the HMC Kernel, Scoring by Expectation, for 3×2 Elections

Label	Time (s)	MLE PHC MSE	Mean PHC MSE	g	t
$d_4 \times v_3$	456.4310	0.1292	0.0922	10	1000
$d_4 \times v_4$	466.3668	0.0492	0.0559	10	1000
$d_5 \times v_3$	479.0481	0.1092	0.1371	10	1000
$d_5 \times v_4$	457.3155	0.0692	0.0486	10	1000
$d_6 \times v_3$	469.6996	0.2025	0.2227	10	1000
$d_6 \times v_4$	475.1872	0.0692	0.0791	10	1000

3.2.6 COMPARING THE CONFIGURATIONS IN THE 2×2 CASE

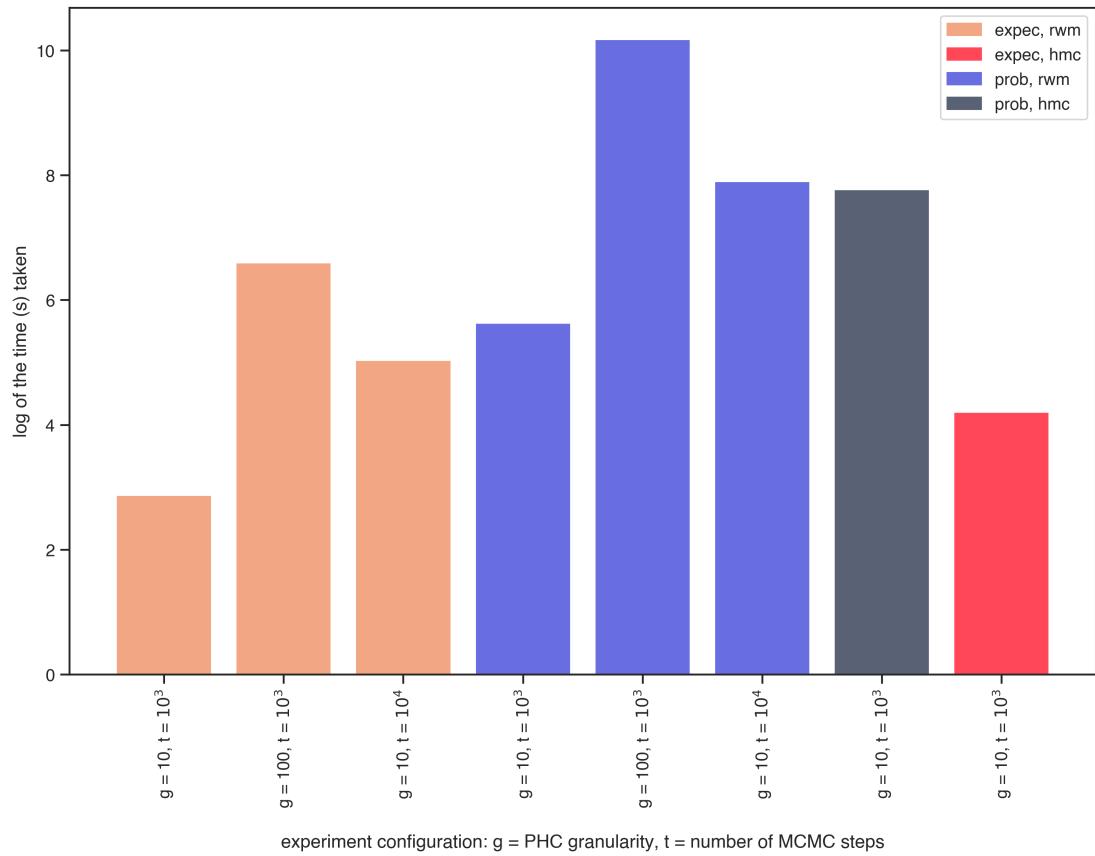


Figure 3.1: Runtime of the Configurations on the $d_1 \times v_1$ Election

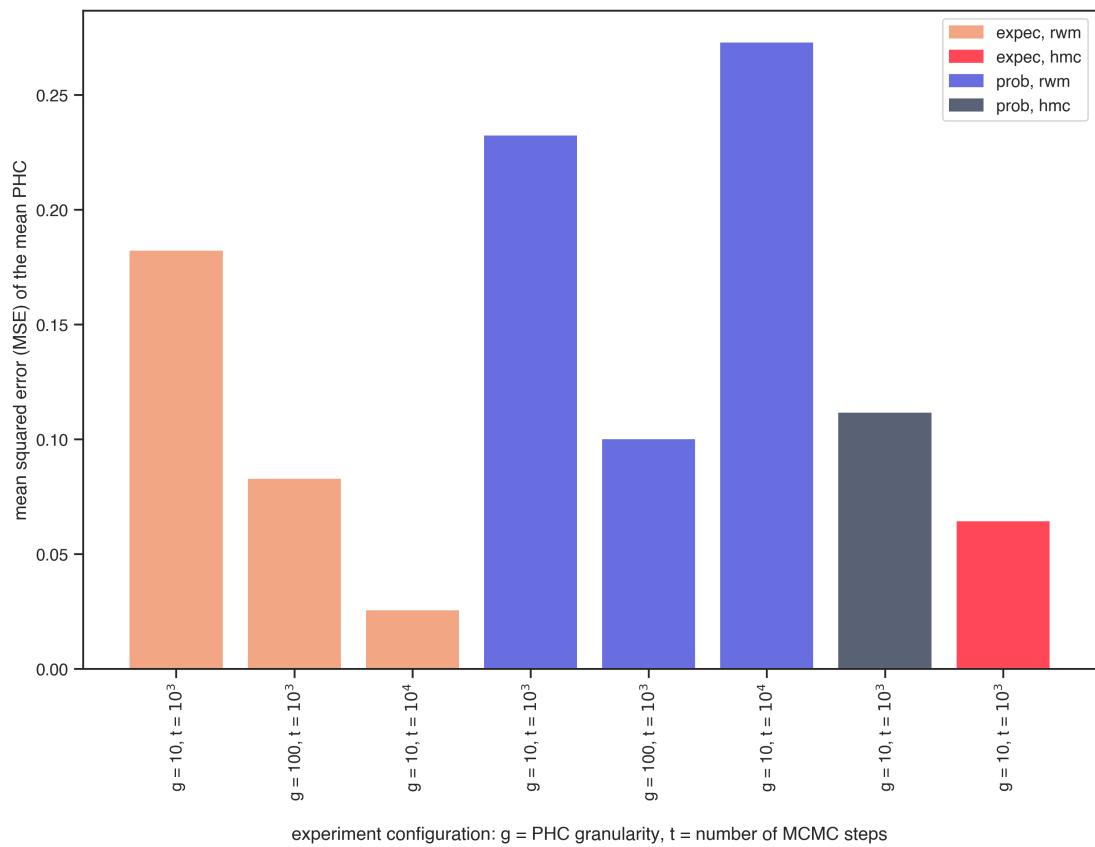


Figure 3.2: MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_1$ Election

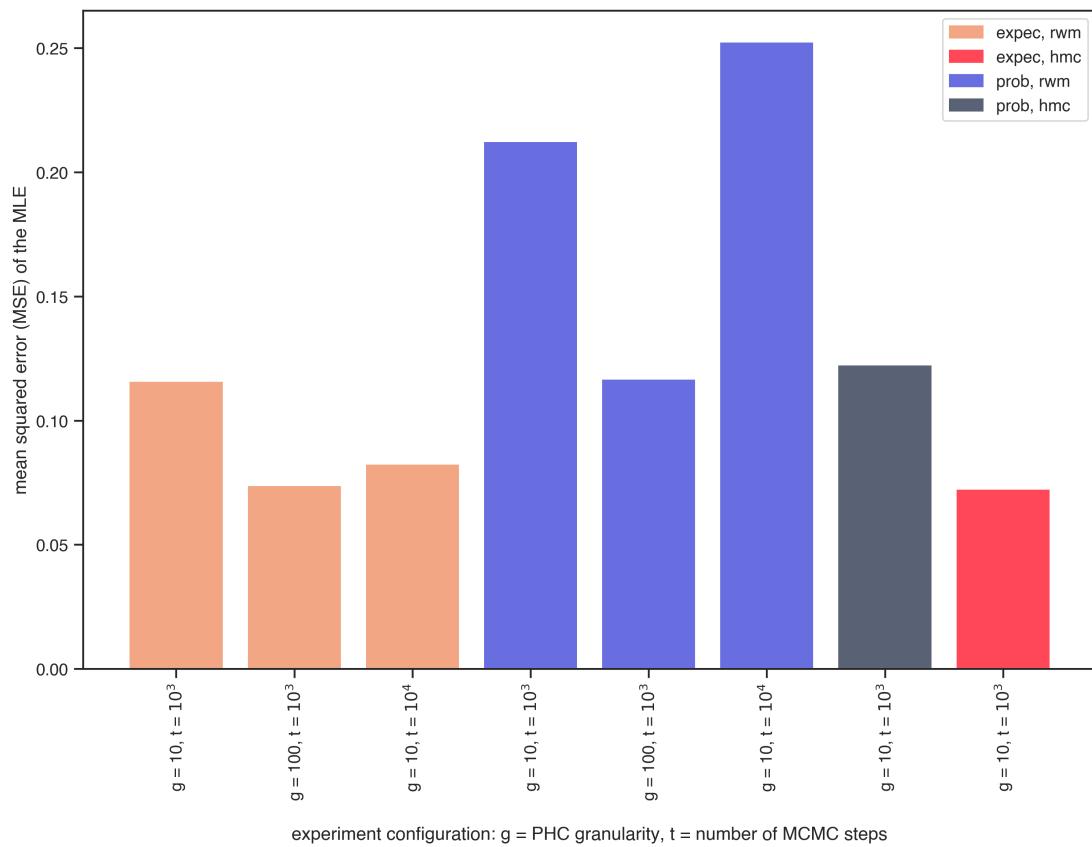


Figure 3.3: MSE of the MLE of the Configurations' DVM on the $d_1 \times v_1$ Election

3.2.7 COMPARING THE CONFIGURATIONS IN THE 3×2 CASE

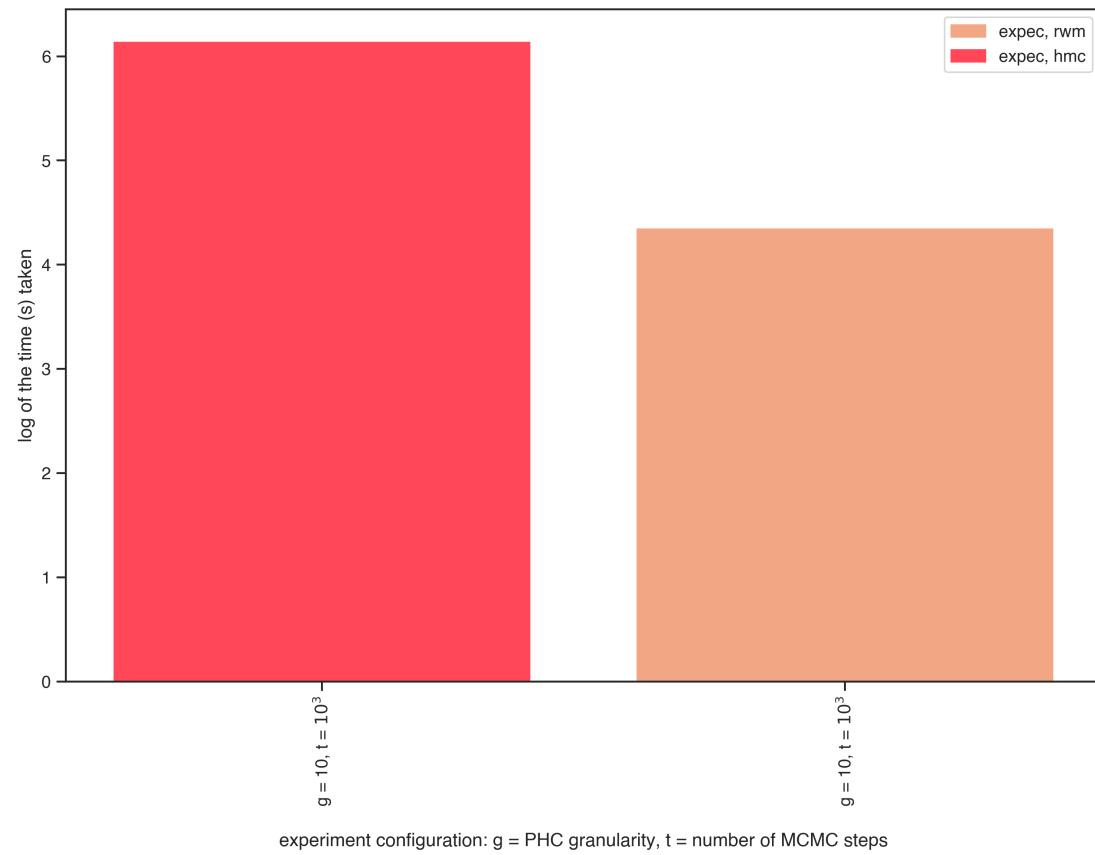


Figure 3.4: Runtime of the Configurations on the $d_4 \times v_4$ Election

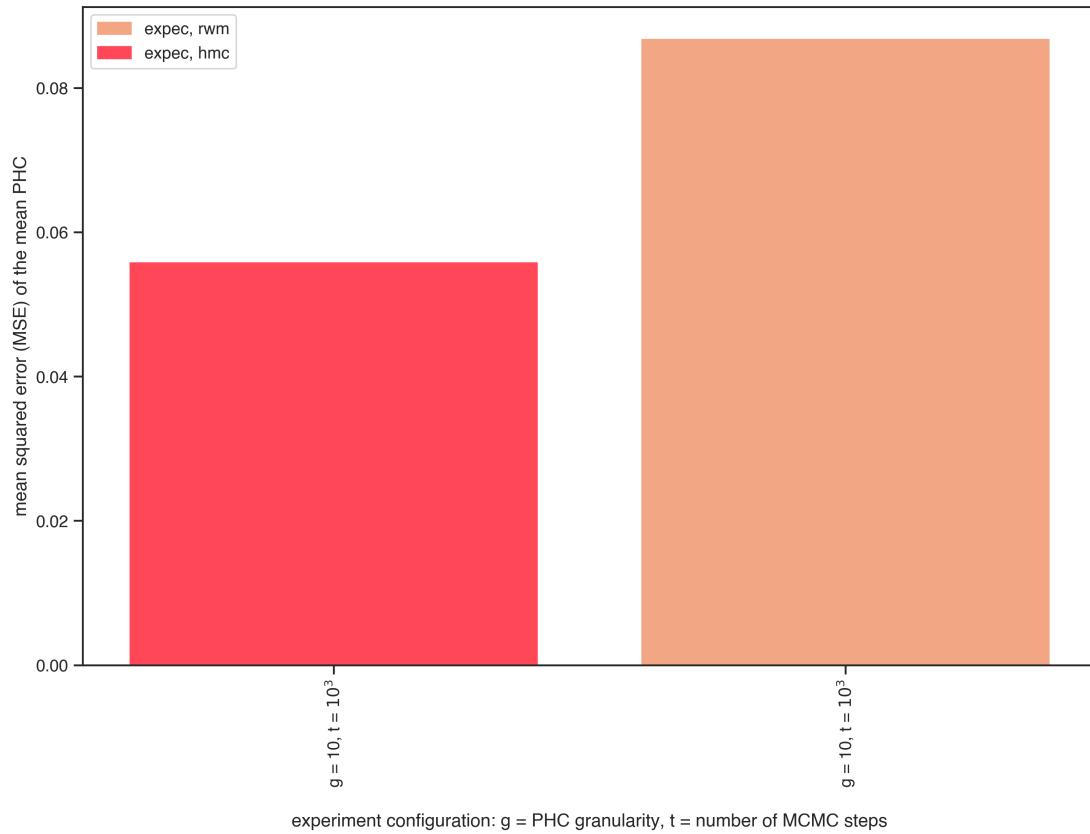


Figure 3.5: MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_4$ Election

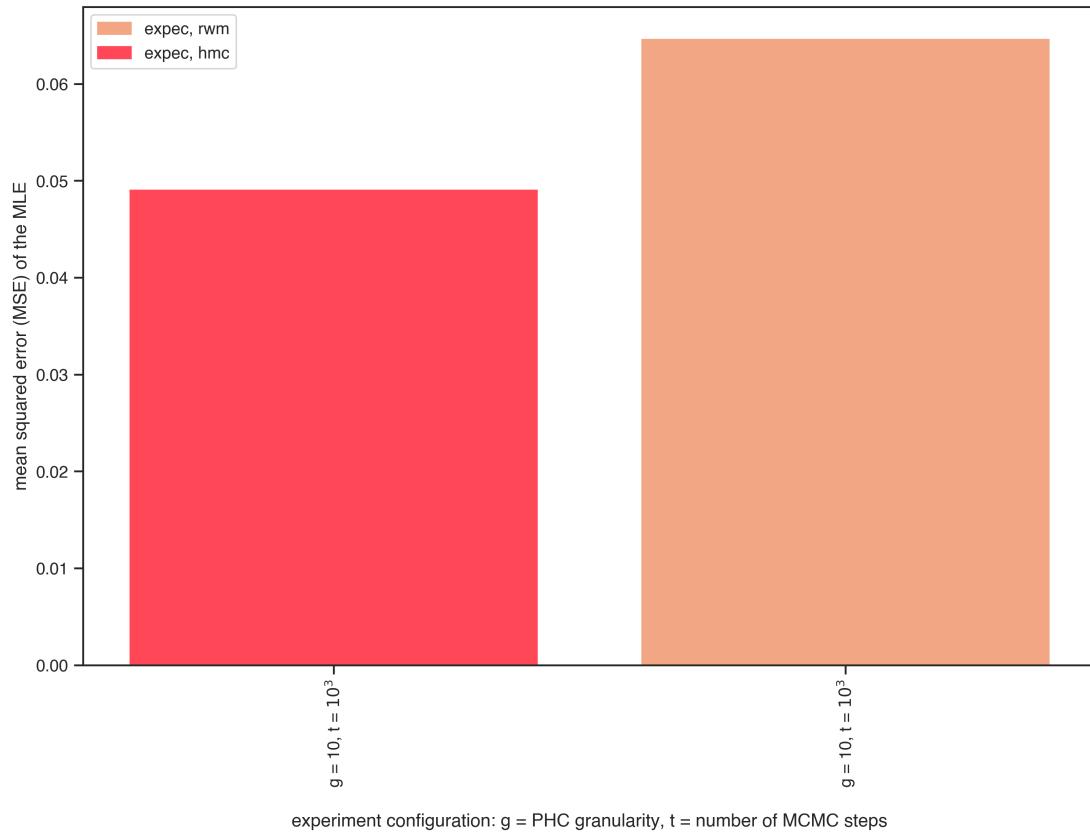


Figure 3.6: MSE of the MLE of the Configurations' DVM on the $d_4 \times v_4$ Election

The work of the political activist inevitably involves a certain tension between the requirement that position be taken on current issues as they arise and the desire that one's contributions will somehow survive the ravages of time.

Dr. Angela Davis

4

Discussion

THE RESULTS OF THE EXPERIMENTS detailed in Chapter 3 suggest a few relationships between the Discrete Voter Model (DVM) and King's Ecological Inference method (KEI):

- i. In the 2×2 case, DVM runs as quickly as KEI for the same number of steps, using Random Walk Metropolis (RWM) as the kernel and expectation-based scoring. It is slower with the

Hamiltonian Monte Carlo kernel and either scoring function, and with RWM and probability-based scoring.

2. In the 2×2 case, KEI is more accurate than DVM in nearly all elections, but the results are close.
3. Unlike KEI, DVM can produce reliable results for the 3×2 case (three demographic groups and two candidates) in a reasonable amount of time.

The results also suggest a few relationships among the DVM configurations. Each of the following relationships is observed in the context of controlling for other variables (*ceteris paribus*: with other conditions remaining the same).

1. The Random Walk Metropolis (RWM) kernel is faster than the Hamiltonian Monte Carlo (HMC) kernel.
2. The HMC kernel produces more accurate results than the RWM kernel.
3. Expectation-based scoring is faster than probability-based scoring.
4. Expectation-based scoring is often just as accurate, if not more accurate, than probability-based scoring.
5. Higher granularity probability hypercubes cause DVM to take more time.
6. When the chain is run for longer, DVM produces more accurate results.

There are many variables contributing to the results of these experiments in more specific ways than the relationships and trends noted above. Section 4.1 discusses their effects.

4.1 THE EFFECT OF DVM'S PARAMETERS

4.1.1 PHC GRANULARITY

THE RESULTS SHOW that the higher the granularity of the probabilistic hypercubes in the model, the slower the model but the more accurate the results.

Figure 3.2 shows that the MSE of the high-granularity PHC is less than half that of the lower-granularity PHC when scoring by either expectation or probability. Figure 3.1 shows that this configuration is the most time-intensive of all configurations with the same kernel and scoring function. Similar trends can be observed in the other elections, illustrated in Tables 3.6 to 3.8 and Figures A.1 to A.12.

With that said, the time-intensiveness of running DVM on such high granularity probability hypercubes made it infeasible to run DVM on elections with more than 2 demographic groups. However, having high granularity PHCs is not the only, and not the best, way to get accurate results in a reasonable time frame on a typical computer.

There is a clear tradeoff between time or computing power and accuracy if one chooses to vary the PHCs' granularities.

4.1.2 NUMBER OF MCMC STEPS

THE RESULTS SHOW that the higher the number of steps in the Markov chain, the more accurate the model, and the slower the model.

Figure 3.2 shows that the MSE of the configuration with more steps was more accurate than the configuration with fewer steps when scoring by expectation. DVM in the configurations with a high number of steps and scored by probability are less accurate than than the lower number of steps and high granularity configurations in election $d_1 \times v_1$. However, in several other elections, like $d_1 \times v_2$ and $d_3 \times v_2$, illustrated in Figures A.8 and A.12 respectively, the configurations with the higher number of steps are more accurate.

Running the chain for more steps is necessarily more time-intensive than running it for fewer, but it is less time-intensive than running for fewer on high granularity PHCs.

Given that running the chain for more steps produces better results in a shorter amount of time than increasing the granularity of the PHCs, it is advisable to increase the number of steps in the chain to increase the accuracy of DVM, instead of increasing the granularity of the PHCs.

4.1.3 KERNEL

THE RESULTS SHOW that the Hamiltonian Monte Carlo (HMC) kernel is more time-intensive, for the same number of steps, than the Random Walk Metropolis (RWM) kernel, but is generally more

accurate.

For most elections in the 2×2 case, including $d_1 \times v_1$ illustrated in Figure 3.2, HMC is more accurate than many of the other configurations. It is most comparable in accuracy to the similarly strong RWM kernel run for 10 times more steps. When comparing the two on the same “number of steps”, however, HMC is more accurate.

This aligns with the research, intuition, and design of HMC. A “step” in HMC is not the same as a “step” in RWM. In the former, the “step” is much more deliberate and intricate, doing several optimizations to move the chain in the direction of highest probability density. Hence, a 1:1 comparison of HMC and RWM, on the basis of steps, is not useful. With that said, one can compare the amount of time that each needs to produce results with a certain amount of accuracy. HMC is consistently on par, if not a little faster, than RWM with a large number of steps. Nearly every figure in the range A.7 to A.12 shows this.

HMC also shines in the 3×2 case, consistently outperforming RWM when both are scored by expectation.

HMC and RWM are comparable when RWM is run for more steps, and when both are run for about the same amount of time. The tradeoff here is the same for deciding to run the chain for longer using RWM: HMC is typically more time- and compute-intensive, but it does produce good results. With that said, RWM’s results are quite comparable, and if time or computing resources are strained, it should probably be preferred.

4.1.4 SCORE FUNCTION

THE RESULTS SHOW that scoring by probability is almost always more time-intensive than scoring by expectation, but does not usually produce more accurate results.

For about half of the elections analyzed, scoring by probability was worse than scoring by expectation. Figures 3.3 and 3.1 are good examples of this. Every expectation-based counterpart to the probability-based configuration was more accurate. The expectation-based configurations were also orders of magnitude faster than the probability-based configurations. The difference is even more stark with HMC configurations, where scoring by expectation was always more accurate than scoring by probability.

In the 3×2 case, like with configurations that increased the number of steps and granularity of the PHCs, scoring by probability was computationally infeasible on the typical computer, within a reasonable time frame.

Hence, scoring by expectation should almost always be preferred when using the Discrete Voter Model.

4.1.5 THE RECOMMENDED CONFIGURATION

LIKE MOST STATISTICAL MODELS, there is no perfect configuration of parameters that will work for all cases. The Discrete Voter Model is implemented such that switching between different configura-

tions of parameters is simple, and that exploration is high encouraged.

With that said, the experiments conducted suggest that using the Random Walk Metropolis kernel, scoring by expectation, and running the model for a large number of steps on moderately-sized probability hypercubes generally produces better samples of the true distribution, and more accurate results.

That configuration was employed on the case studies from Chicago presented in Section 4.2.

4.2 CASE STUDIES: CHICAGO 2015 AND 2019 MAYORAL RUNOFFS

CHICAGO is a large, beautiful and diverse city with a strong political system and culture of civic participation. Chicago also suffers from well-known problems in its electoral and civic structures, including gentrification, racial segregation, and gerrymandering.

The Metric Geometry and Gerrymandering Group (MGGG) notes in their *Study of Reform Proposals for Chicago City Council*:

The wards are perceived to be gerrymandered, with eccentric boundaries that preserve some traditionally favored neighborhoods but wind awkwardly through other parts of the city, denying Austin and Little Village the political stability that Bridgeport and Beverly have enjoyed for years. They also tend toward extreme malapportionment by the end of their ten-year cycle, with the largest ward having twice the population of the smallest, flouting the Constitutionally recognized principle of “One Person, One Vote.” Further, we will demonstrate that the current wards are massively racially segregated, far beyond the levels that would be expected in a 50-ward plan, and exhibit high levels of economic stratification as well. And legacy policies like “aldermanic privilege” combine with carefully chosen lines to entrench some aldermen beyond the reach of

accountability to their voters, creating what some have described as a system of fifty fiefdoms.²²

This damning assessment of Chicago's electoral system, and particularly the suspected gerrymandering of its wards, makes it an excellent case study for the Discrete Voter Model.

Of the 2069 voting precincts in Chicago, 563 are more than 90% Black while 1054 have less than 10% Black population.²² To help visualize the dramatic racial segregation, Figure 4.1 shows the chloropleth maps of the four largest racial groups in Chicago. Similar patterns exist along other axes of identity, like socioeconomic status.

Chicago's diverse and politically active electorate has more than the 2 demographic groups that previous methods of determining racially polarized voting would allow. That racial diversity, combined with the massive amounts of segregation, sets up the perfect storm for racial gerrymandering, and necessitates a method like the Discrete Voter Model. The Discrete Voter Model also does not need to use race as the category for its demographic groups. Models can be easily built along other separations of the electorate, like socioeconomic status and age, or combinations of them.

Two recent runoff Mayoral elections, between Rahm Emanuel and Jesús "Chuy" García in 2015, and Lori Lightfoot and Toni Preckwinkle in 2019, are analyzed in this case study, with race as the demographic groups of interest.

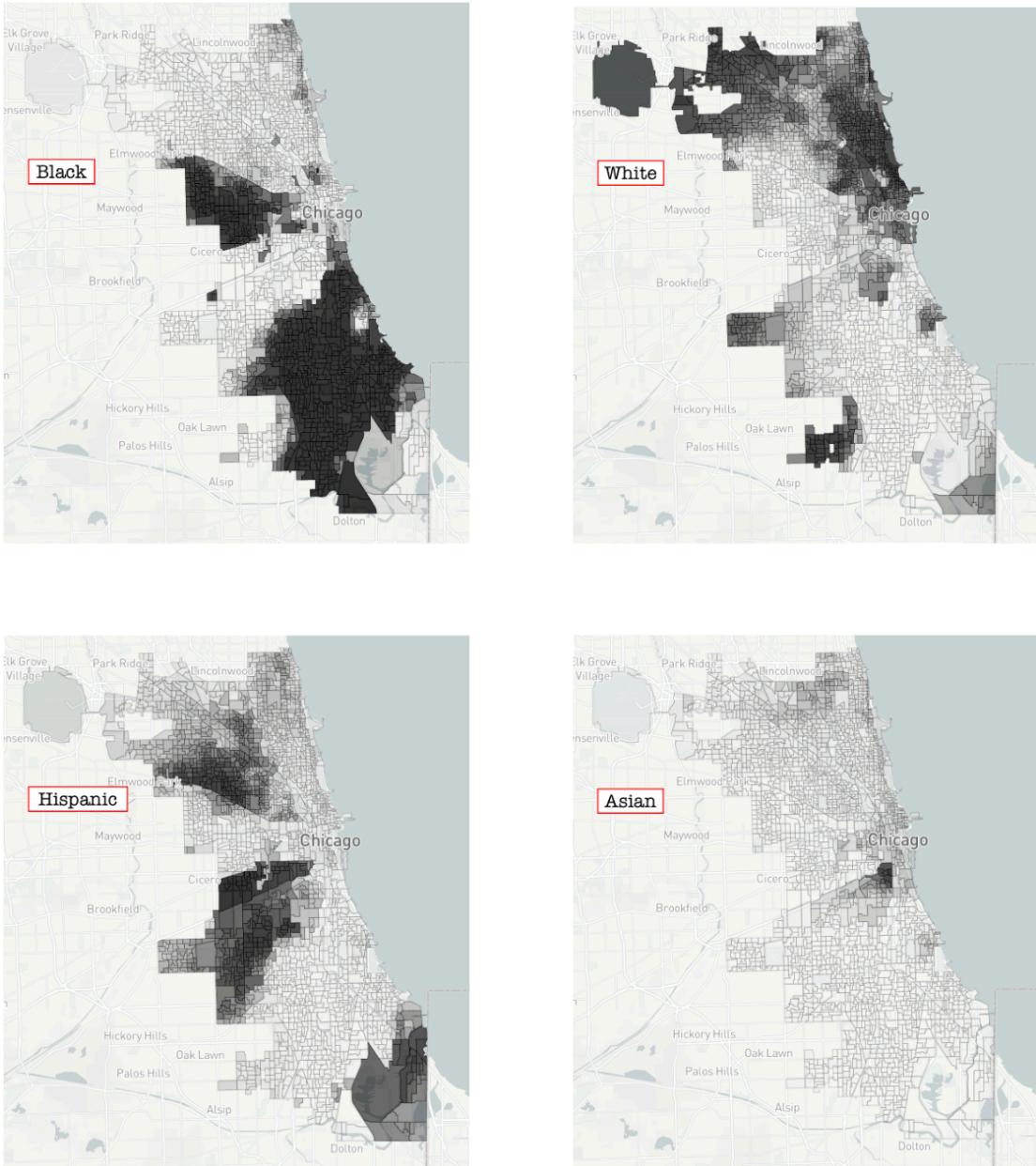


Figure 4.1: Choropleth Maps of the Four Largest Racial Groups in Chicago. Figure 2 in MGGG's Chicago Report.²²

4.2.1 DATA GATHERING

ELECTORAL AND DEMOGRAPHIC DATA for Chicago were generously provided by the [Metric Geometry and Gerrymandering Group](#), and is available publicly in a [GitHub repository](#).

The data were cleaned to extract the necessary fields for the Discrete Voter Model, namely the precinct IDs, ward IDs, demographic fields, and electoral outcome fields.

In order to facilitate visualization, the data presented here were also clipped to include the three racial groups with the highest populations, identified as they are in the Census: Black, white, and Hispanic Chicagoans. However, model was still successfully run on full versions of the electorate, which also included Asian and other Chicagoans.

4.2.2 RESULTS

THE DISCRETE VOTER MODEL was applied to Chicago's 2015 and 2019 Mayoral runoff elections, with Black, white, and Hispanic voters as the demographic groups.

DVM was configured to run for 10000 steps, with the Random Walk Metropolis kernel and expectation-based scoring on granularity 10 probability hypercubes, with a burn-in fraction of 0.3.

Figures 4.2 and 4.3 show the trace plots of the Markov chain. The higher the point on the y axis, the more likely the probabilistic hypercube at that state was to have been able to generate the outcome of the election. A trace plot of a well-mixed and converged Markov chain oscillates a bit, then plateaus

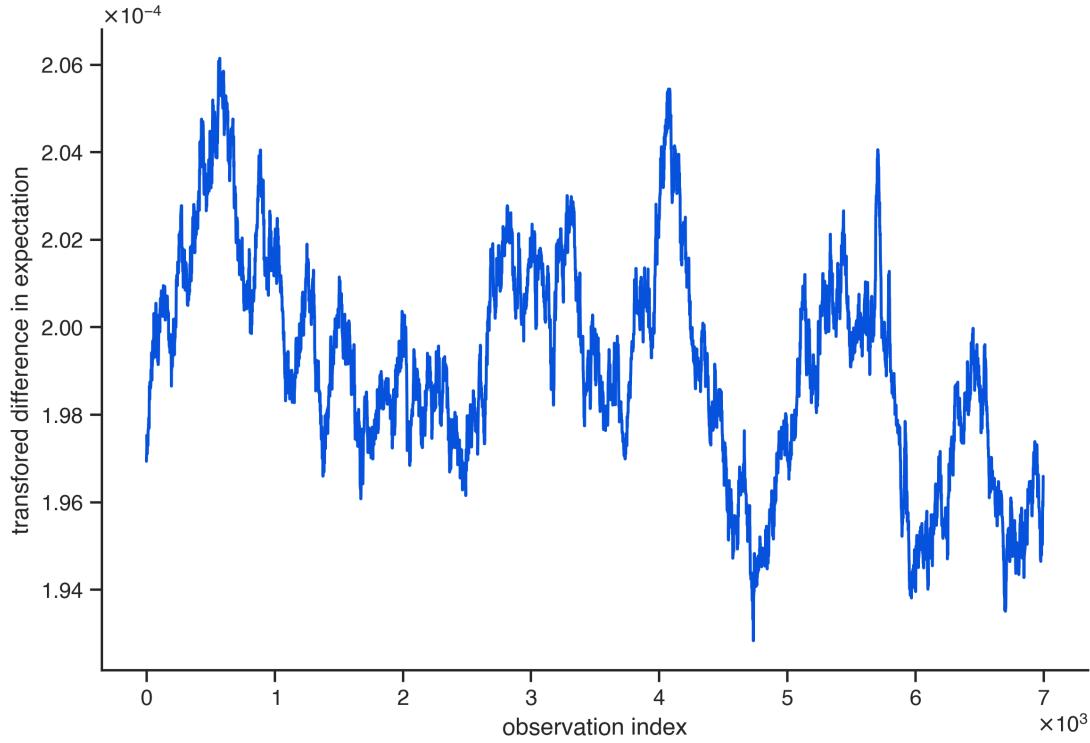


Figure 4.2: A Trace of the Sigmoid Difference in Expectation of the Probabilistic Hypercubes in the DVM Chain for Chicago's 2015 Mayoral Runoff Election

near the end of the sample. Neither of these trace plots show particularly poor behavior, but it is not clear that the chain has converged, or is close to convergence, either.

Figures 4.4 to 4.7 illustrate the mean and MLE probabilistic hypercubes for the sample generated by DVM. They are all quite intuitive in their suggestions. In neither election is much probability concentrated at the edges of the PHC. If that were the case, there would be massive amounts of racially polarized voting. Instead, the PHCs show some probability concentrated at the edges, with most of it toward the center.

The concentration of probability at the edges suggests that there is at least some racial voting polar-

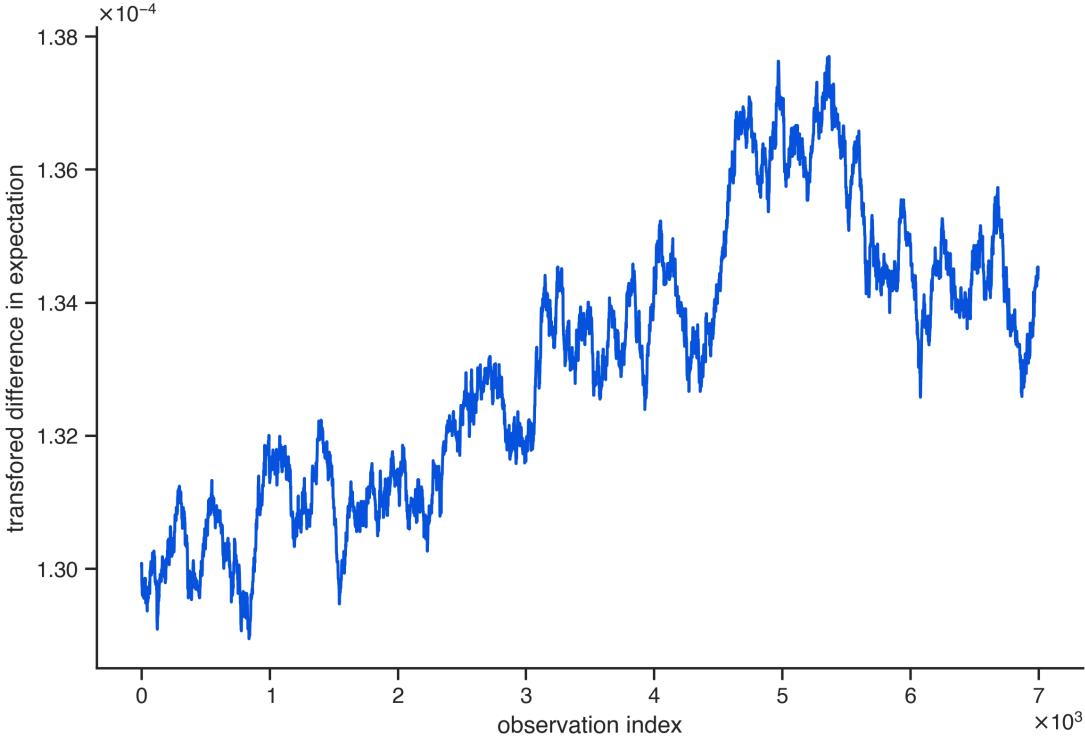


Figure 4.3: A Trace of the Sigmoid Difference in Expectation of the Probabilistic Hypercubes in the DVM Chain for Chicago's 2019 Mayoral Runoff Election

ization, but that it is not unreasonably dominant. For example, Figure 4.6, the mean PHC for Lori Lightfoot in the 2019 Chicago Mayoral runoff election, shows that there is a relatively high probability that she was quite popular with Hispanic and Black voters, which does match anecdotal evidence.⁸.

These probabilistic hypercubes can then be used with precinct data to choose some demographic voting probability. For example, using the PHC represented in Figure 4.6, one can supply a ward, say ward 1, and find that a likely set of demographic voting probabilities for that ward is that $\approx 90\%$ of Black voters, $\approx 55\%$ of Hispanic voters, and $\approx 25\%$ of white voters preferred Lori Lightfoot over Toni Preckwinkle. Those numbers, inferred with the Discrete Voter Model, are the unknowns in

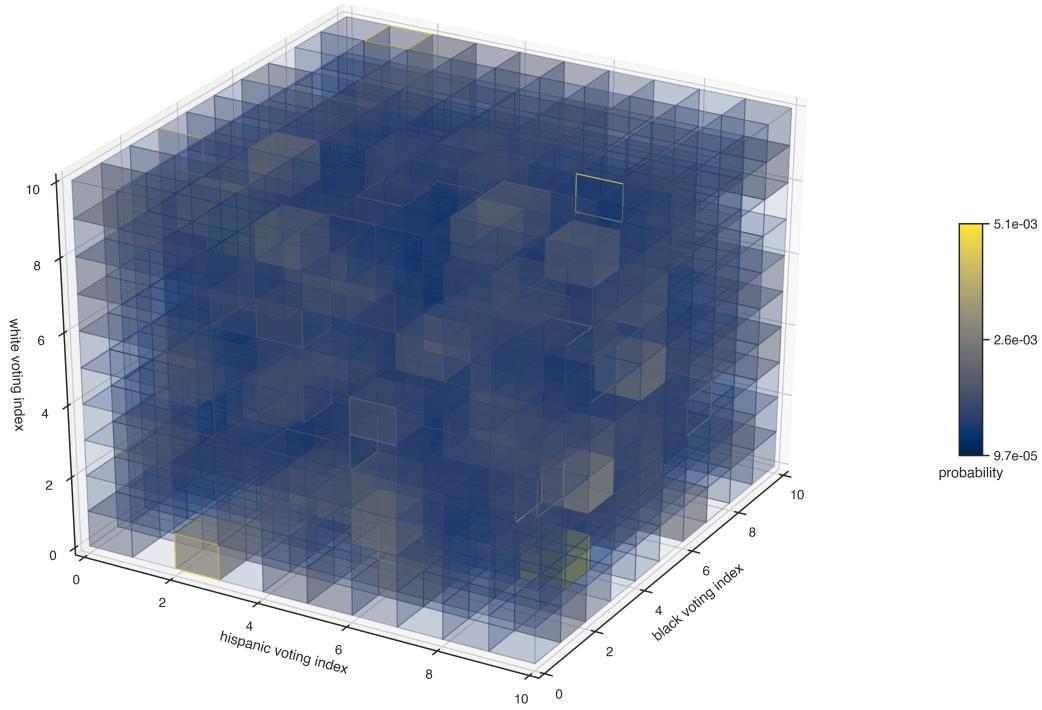


Figure 4.4: The Mean PHC for the 2015 Chicago Mayoral Runoff, for Rahm Emanuel

the ecological inference problem. Black and Hispanic voters seem to prefer this candidate, so there is some racial voting polarization. While these results are the outcome of a probabilistic process, and thus inherently non-deterministic, the experiment and case study are quite repeatable. Not only do they yield similar results, but they also yield similar probability hypercube distributions. The PHC object can thus express more about the Chicagoan electoral landscape than ever before.

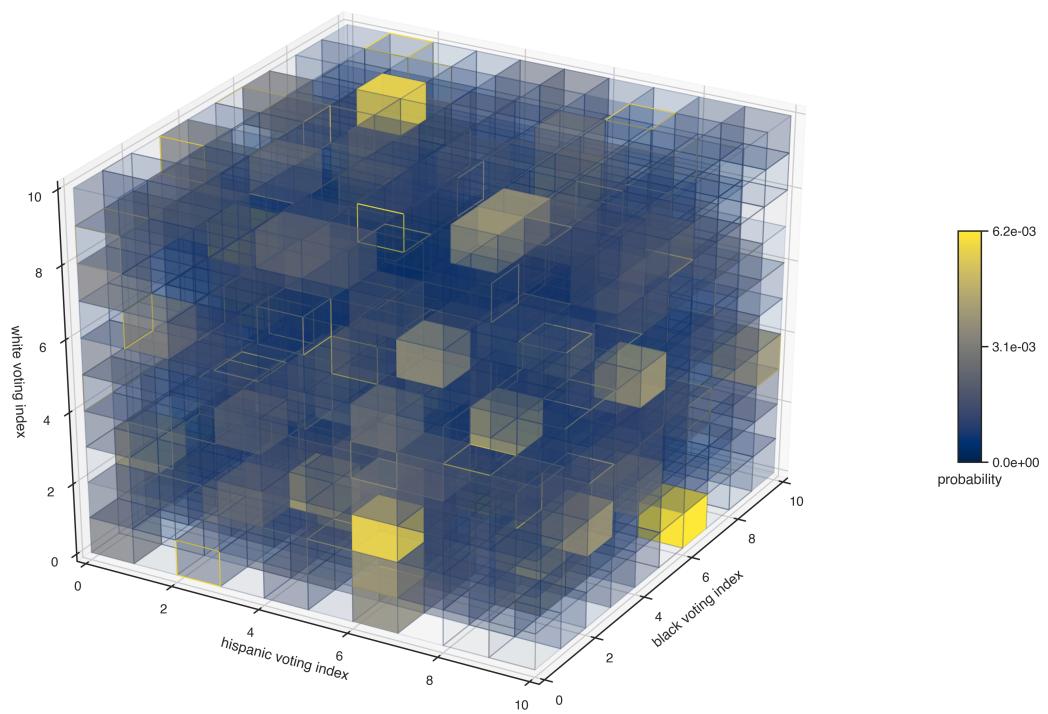


Figure 4.5: The MLE for the 2015 Chicago Mayoral Runoff, for Rahm Emanuel

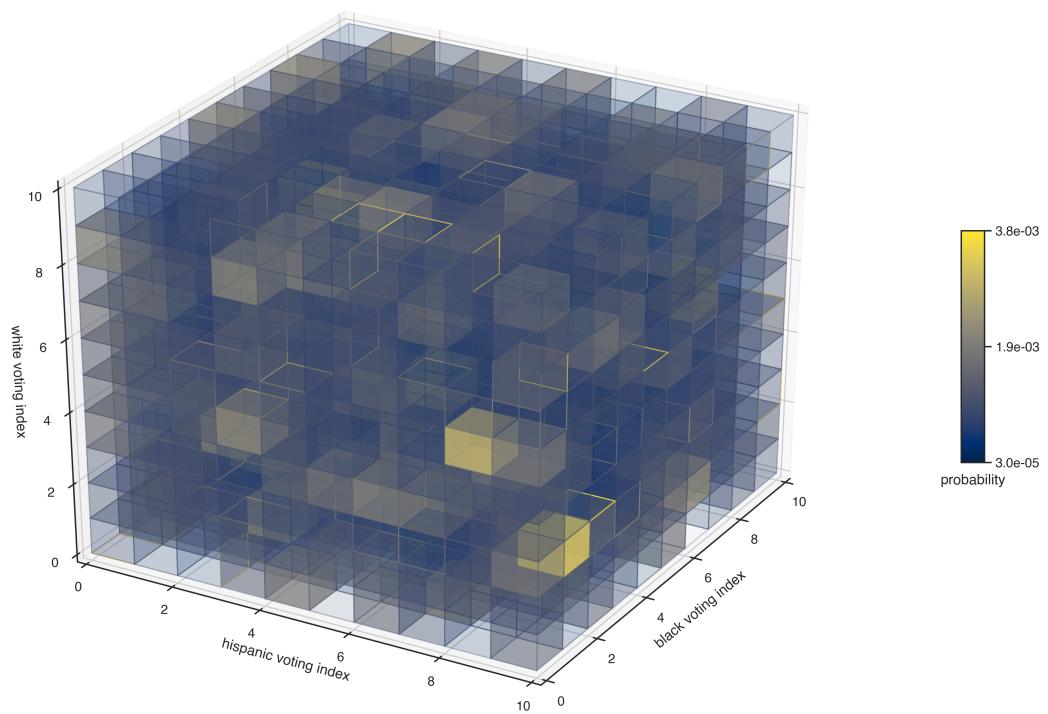


Figure 4.6: The Mean PHC for the 2019 Chicago Mayoral Runoff, for Lori Lightfoot

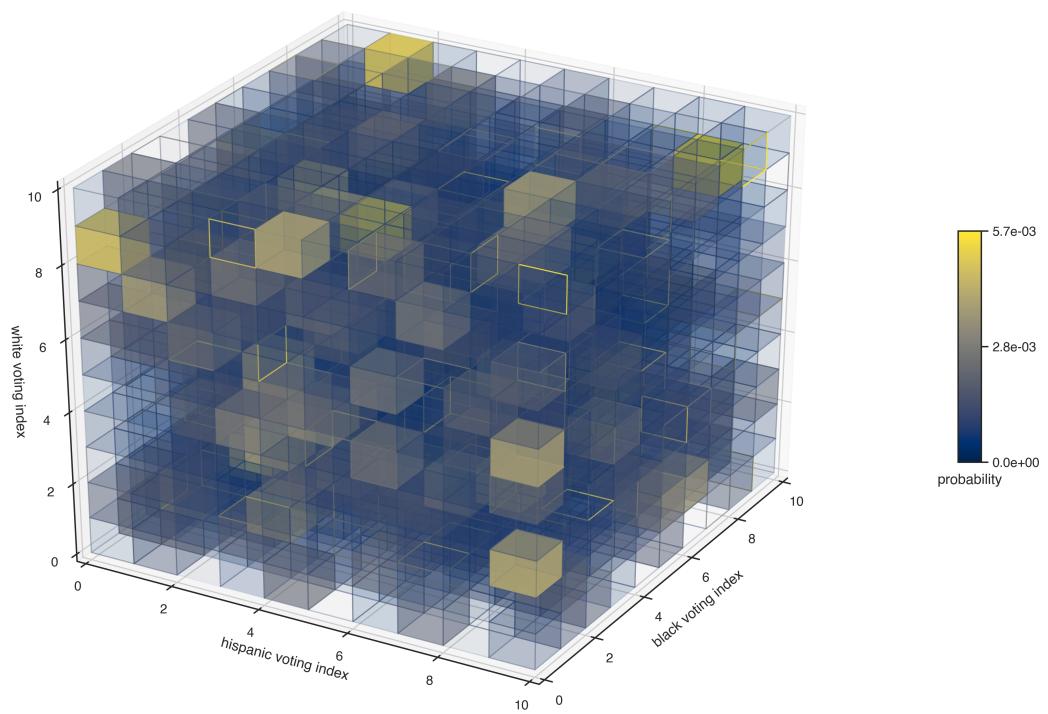


Figure 4.7: The MLE for the 2019 Chicago Mayoral Runoff, for Lori Lightfoot

4.3 IMPLICATIONS

THE STRENGTH OF THE DISCRETE VOTER MODEL is its ability to represent the complex voting relationships between and within demographic groups, no matter how many groups there are. The only limitation is time and computing power, but DVM is already written such that it can be easily scaled to more powerful modern architectures. The rise of cloud computing, and particularly infrastructure as a service (IaaS) products like Google Compute Engine and AWS EC2, allows for the relatively inexpensive rental of unprecedented computing power. The experiments in Chapter 3 and the case study in Section 4.2 were run on a typical laptop computer, and they were still able to produce good results.

The Discrete Voter Model's flexibility will allow for the analysis of how a bevy of important aspects of voters' identities and experiences affect their ability to vote and that vote's power in elections. While race is unfortunately a critically important determinant of one's political power, other characteristics, like age and income, may also affect that power. Since the "demographic group" is abstract, it can be easily replaced to analyze these other characteristics. DVM not only scales with the number demographic groups, but also the size of the election. Hence, these methods can be applied to cities like Chicago, and states like North Carolina without changing the statistical foundation or the implementation.

4.4 FUTURE WORK

THE DISCRETE VOTER MODEL will continue to be evaluated on more complex elections, for longer runs, and on more powerful computing infrastructure. It will also be continually maintained and improved. Options for future work include:

1. adding more transition kernel options
2. adding more scoring function options
3. translating the model's execution phase to a more numerically-performant language, like C++
4. exploring other shapes for probabilistic hypercubes. Promising alternatives include making the n dimensions more abstract, and working only with flat tensors to improve performance.

5

Conclusion

The Discrete Voter Model (DVM) is a novel statistical method for performing ecological inference.

The center of this method is the probabilistic hypercube (PHC), a rich data structure that encodes complex voting patterns in elections and electorates.

DVM leverages the power of discretization and Markov chain Monte Carlo to provide more expressive models within a more malleable and intuitive structure. In lower dimensions, the results of

DVM can be easily visualized to provide more information about the distribution of possibilities of individual-level behavior than point estimates ever could. Even in higher dimensions, where visualization is not possible, the results of DVM are still quite expressive, and the distribution can be used to discover more about the data.

While DVM was developed for use on elections, it can be thought of and applied more foundationally as a tool for all kinds of ecological inference. If one would like to infer individual-level behavior when only aggregate group data is available, DVM can do that.

DVM is also built on state-of-the-art computing paradigms, like the dataflow and probabilistic programming provided by TensorFlow Probability. This ensures that the model, and the algorithms it implements, continues to leverage novel technologies and improvements in computing architecture.

The development of DVM was inspired by the pressing need to address the chronic problem of racial gerrymandering. Everyone, regardless of race or other personal identifiers, ought to be able to exercise their right to vote, and have that vote count towards change. A country will never be a true democracy until all barriers to voting, including more insidious and systemic ones like a gerrymandered electoral map, are destroyed. Hopefully, by helping to identify egregious racially polarized voting across the country, the Discrete Voter Model contributes positively to that effort.

A

Appendix

The code for the Discrete Voter Model can be found in full in [this GitHub repository](#).

That repository also contains notebooks to run and test the modules of DVM. To completely reproduce these experiments, run `mock_experiments.ipynb` and `case_experiments` in full.

The following are notable code snippets mentioned in the body of the article.

Listing A.1: Python Implementation of King's Ecological Inference

```
import numpy as np

import pymc3 as pm

def eco_inf(prec_demos, first_cand_obs_votes, lmbda=0.5):

    """
    Run King's Ecological Inference method for the 2x2 case
    (2 demographic groups, and 2 candidates).

    prec_demos (list of dicts): the demographics of the precincts
    first_cand_obs_votes (NumPy array): the number of people in each
    precinct who voted for the first candidate
    lmbda (float): the hyperparameter for the Exponential distributions

    return: the probabilistic model
    """

    # Convert the demographics of the precincts to a NumPy array
    total_pop = np.array([sum(demo.values()) for demo in prec_demos])

    # Find the percentage of people in the first demographic group in the
    # precincts
    first_group = list(prec_demos[0].keys())[0]
```

```

demo_pcts = np.array([demo[first_group] for demo in prec_demos]) /
    total_pop

# Find the number of precincts

p = total_pop.size

with pm.Model() as model:

    c_1 = pm.Exponential('c_1', lmbda)

    d_1 = pm.Exponential('d_1', lmbda)

    c_2 = pm.Exponential('c_2', lmbda)

    d_2 = pm.Exponential('d_2', lmbda)

    b_1 = pm.Beta('b_1', alpha=c_1, beta=d_1, shape=p)

    b_2 = pm.Beta('b_2', alpha=c_2, beta=d_2, shape=p)

    theta = demo_pcts * b_1 + (1 - demo_pcts) * b_2

    Tprime = pm.Binomial('Tprime', n=total_pop, p=theta, observed=
        first_cand_obs_votes)

return model

```

Listing A.2: Integer Partitioning

```
from itertools import chain, permutations

def integer_partition(n, k, min_size=0):
    """
    Partition an integer.

    n (int): the integer to partition
    k (int): the number of elements in a partition
    min_size (int): the minimum size of an element
    in the partition

    return: a generator of partitions as tuples
    """

    if k < 1:
        return

    if k == 1:
        if n >= min_size:
            yield (n,)

    return
    for i in range(min_size, n // k + 1):
        for result in integer_partition(n - i, k - 1, i):
```

```
        yield (i,) + result

def permute_integer_partition(n, k, min_size=0):
    """
    Partition an integer, with all permutations

    n (int): the integer to partition
    k (int): the number of elements in a partition
    min_size (int): the minimum size of an element
    in the partition

    return: a generator of all permutations of partitions as tuples
    """
    return chain.from_iterable(set(permute_integer_partition(n - p, k - 1, min_size)) for p in
                               integer_partition(n, k, min_size))
```

A.I BAR PLOTS COMPARING EXPERIMENT RESULTS FOR ALL ELECTIONS

A.I.I RUNTIME OF 2×2 CASES

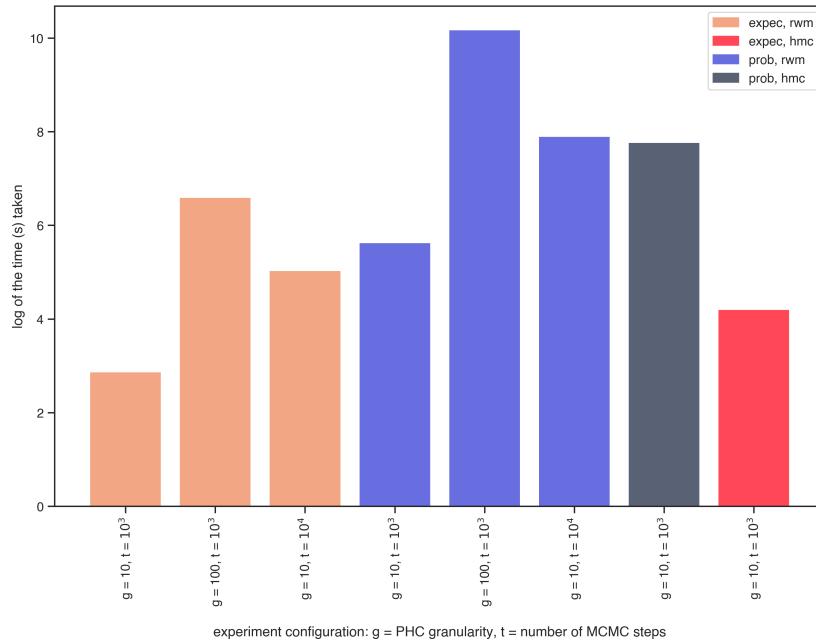


Figure A.1: Runtime of the Configurations on the $d_1 \times v_1$ Election

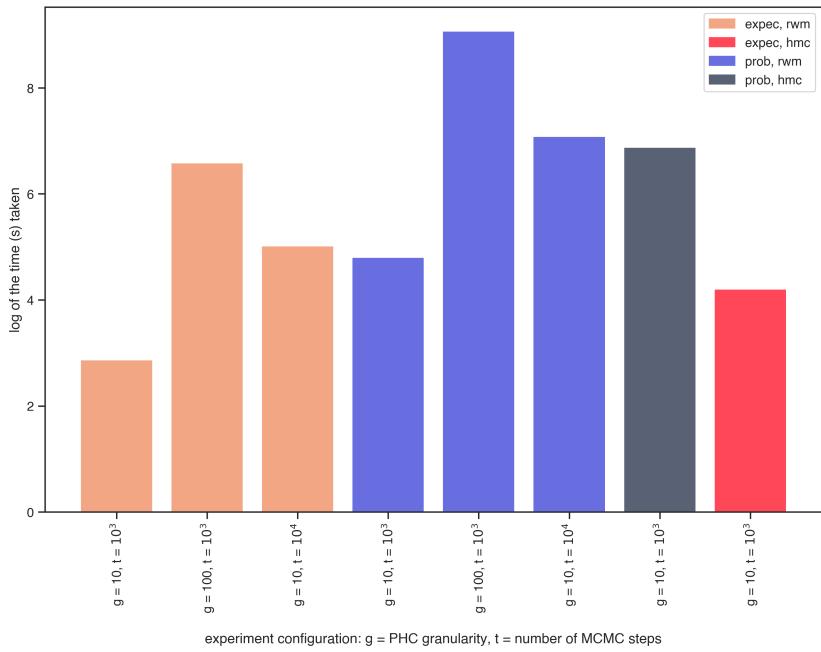


Figure A.2: Runtime of the Configurations on the $d_1 \times v_2$ Election

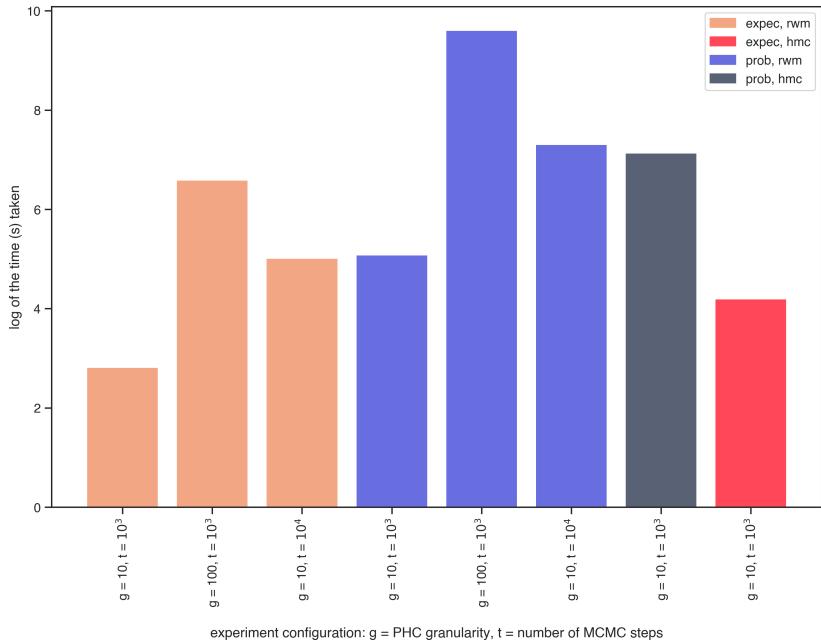


Figure A.3: Runtime of the Configurations on the $d_2 \times v_1$ Election

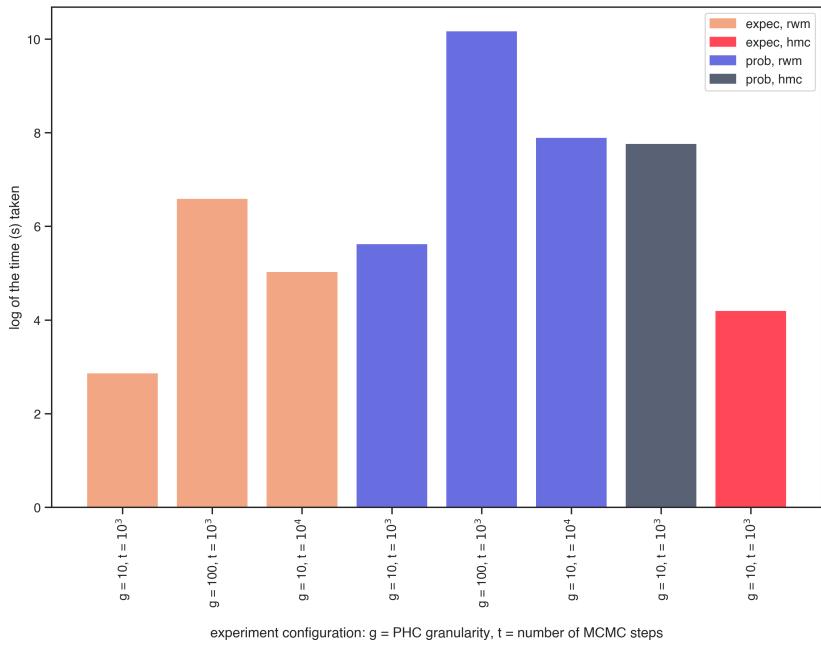


Figure A.4: Runtime of the Configurations on the $d_2 \times v_2$ Election

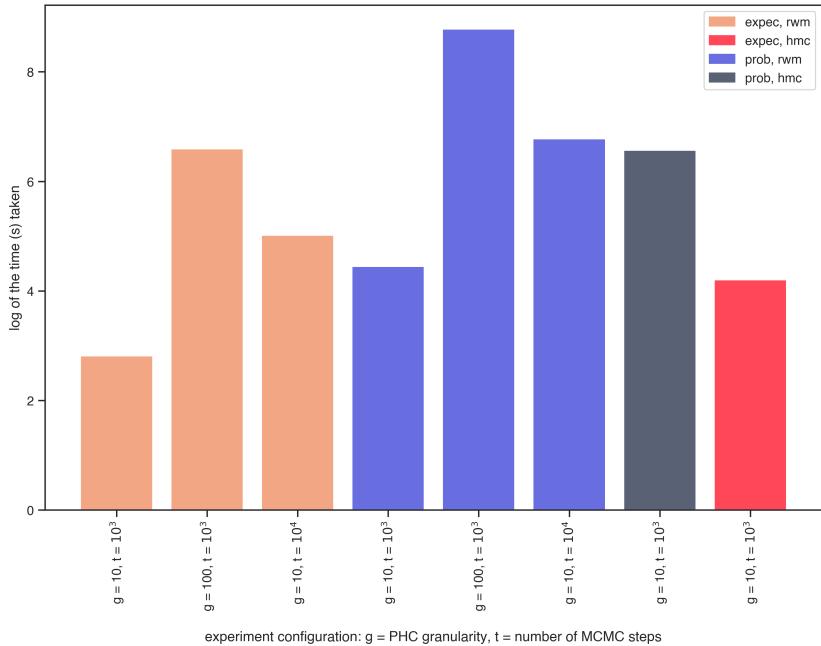


Figure A.5: Runtime of the Configurations on the $d_3 \times v_1$ Election

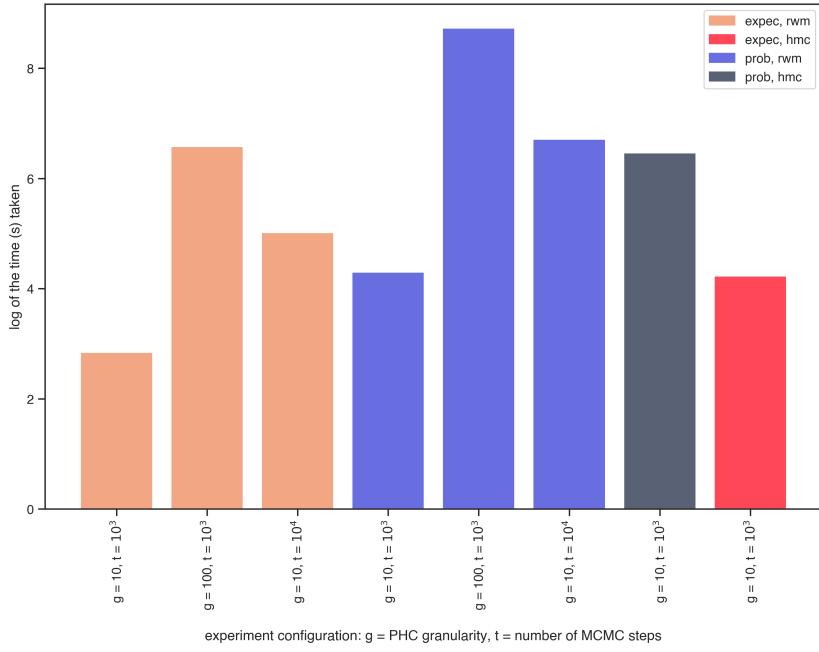


Figure A.6: Runtime of the Configurations on the $d_3 \times v_2$ Election

A.I.2 MEAN PHC MSE OF 2×2 CASES

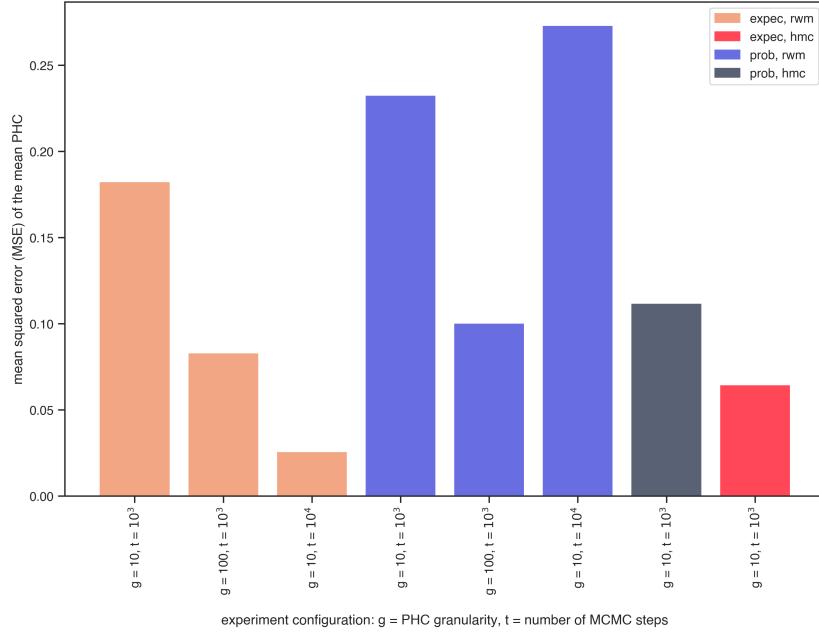


Figure A.7: MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_1$ Election

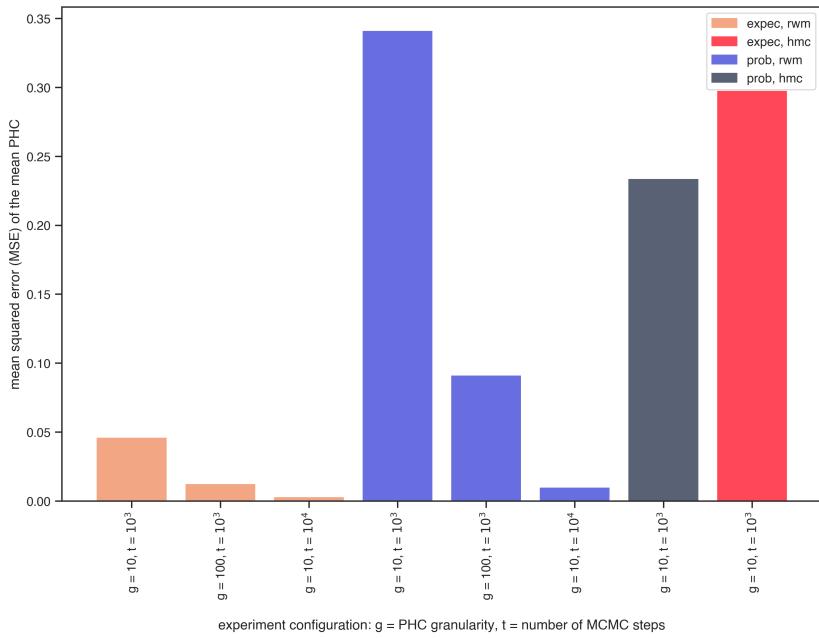


Figure A.8: MSE of the Mean PHC of the Configurations' DVM on the $d_1 \times v_2$ Election

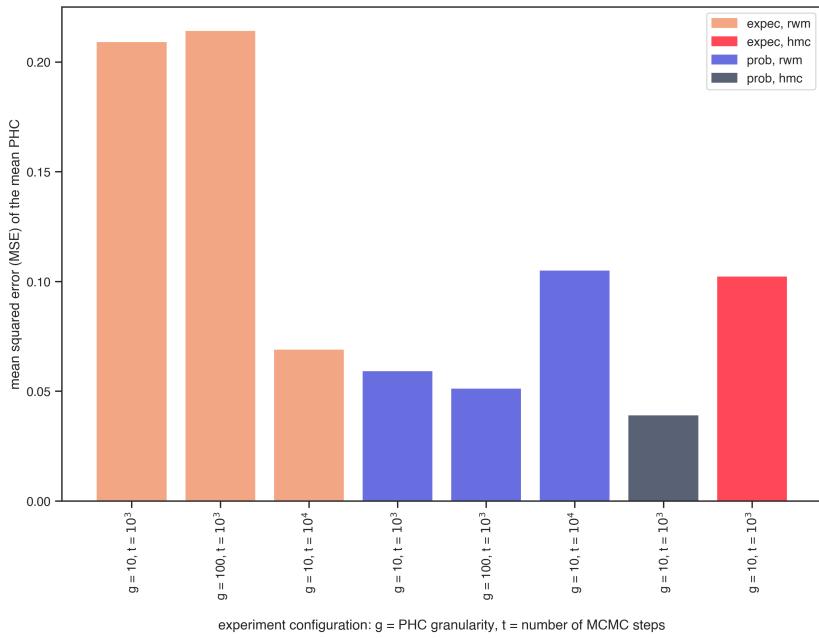


Figure A.9: MSE of the Mean PHC of the Configurations' DVM on the $d_2 \times v_1$ Election

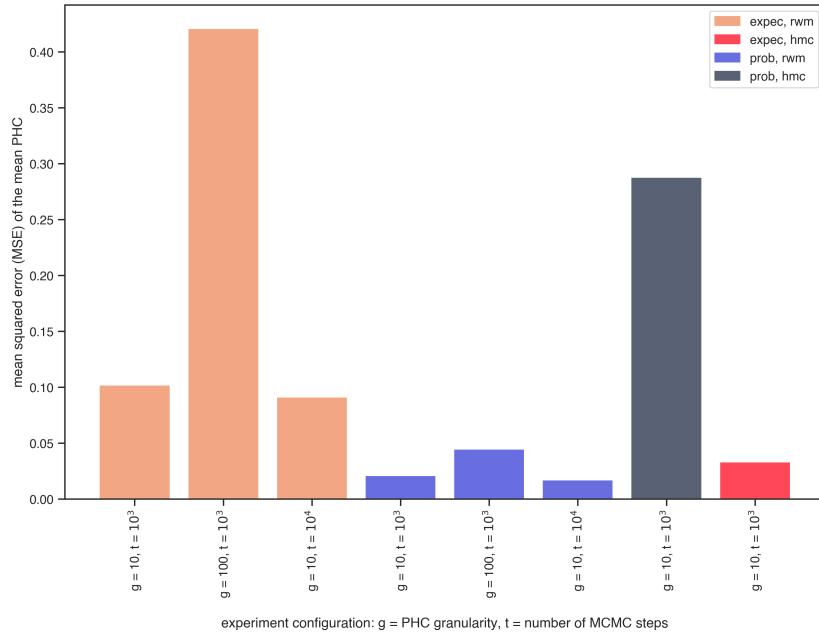


Figure A.10: MSE of the Mean PHC of the Configurations' DVM on the $d_2 \times v_2$ Election

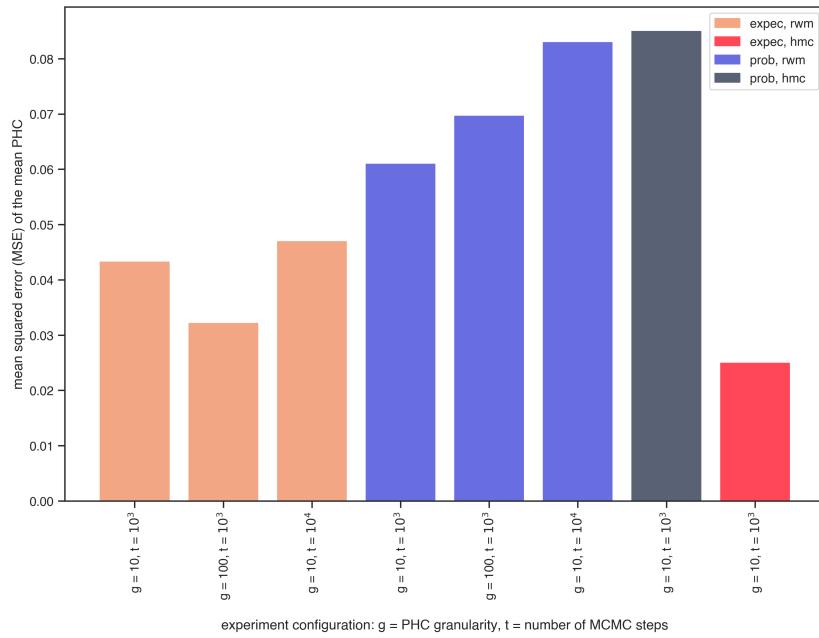


Figure A.11: MSE of the Mean PHC of the Configurations' DVM on the $d_3 \times v_1$ Election

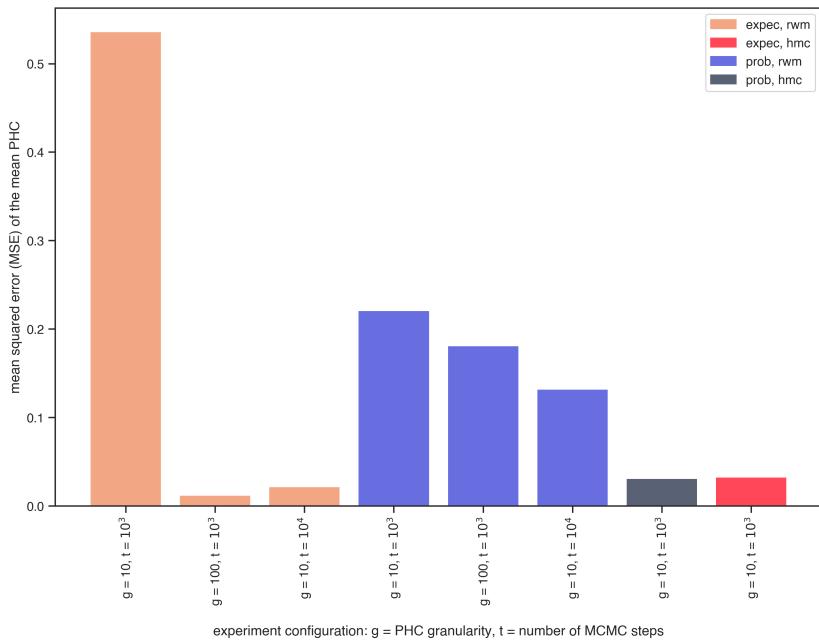


Figure A.12: MSE of the Mean PHC of the Configurations' DVM on the $d_3 \times v_2$ Election

A.1.3 RUNTIME OF 3×2 CASES

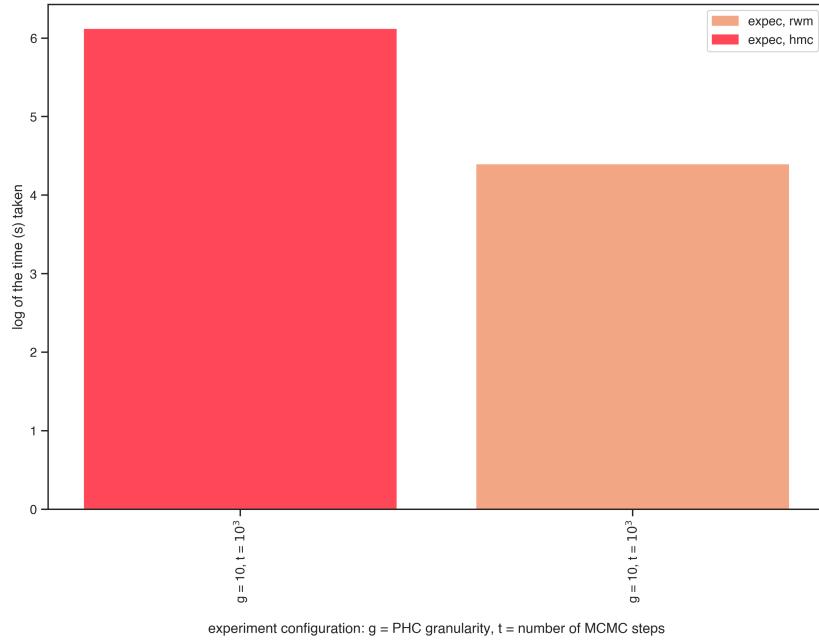


Figure A.13: Runtime of the Configurations on the $d_4 \times v_3$ Election

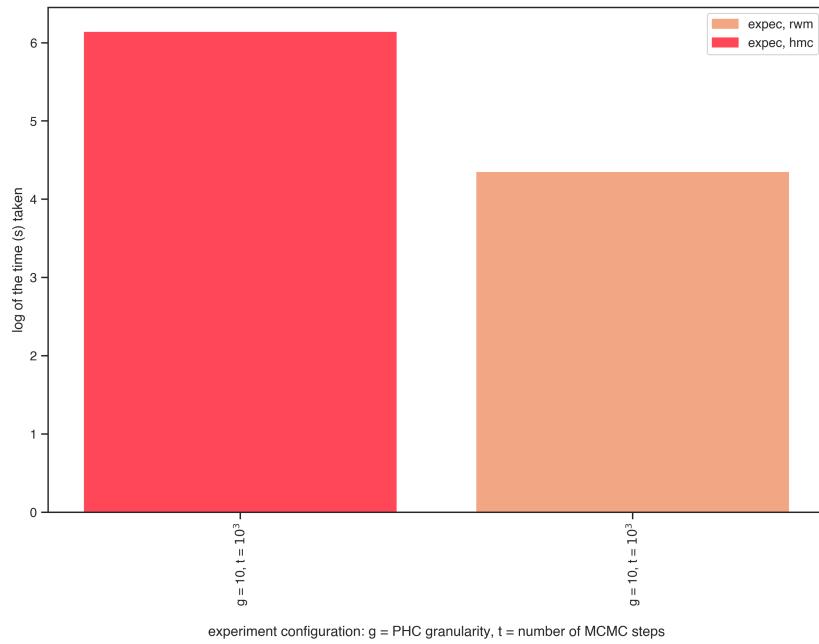


Figure A.14: Runtime of the Configurations on the $d_4 \times v_4$ Election

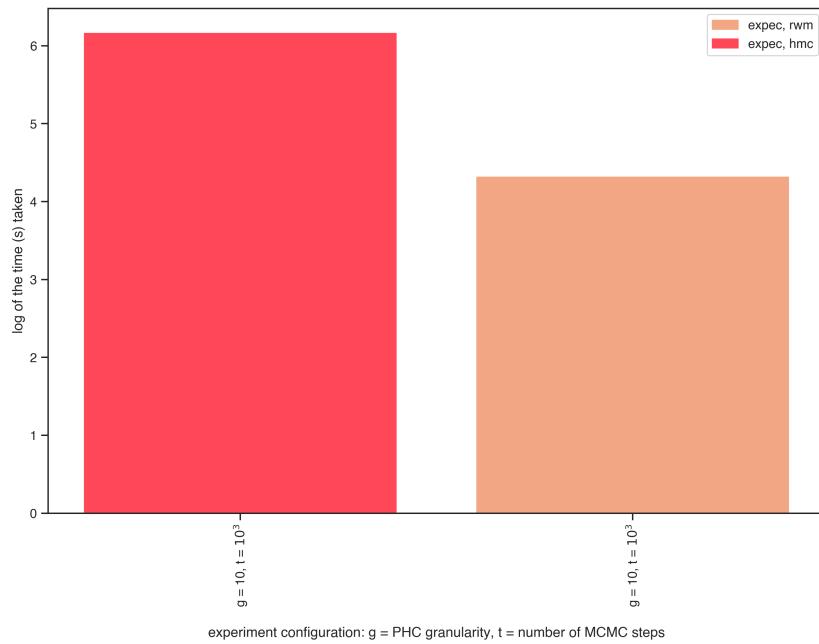


Figure A.15: Runtime of the Configurations on the $d_5 \times v_3$ Election

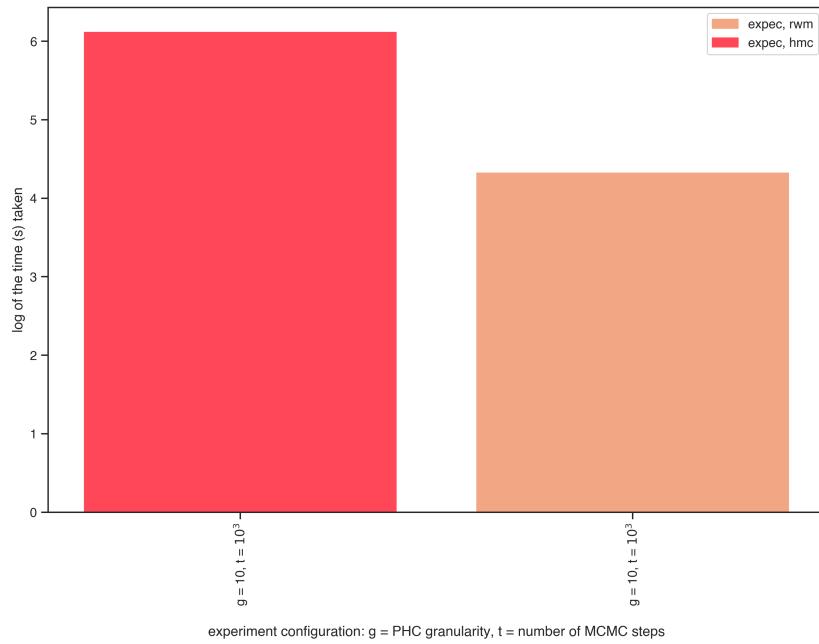


Figure A.16: Runtime of the Configurations on the $d_5 \times v_4$ Election

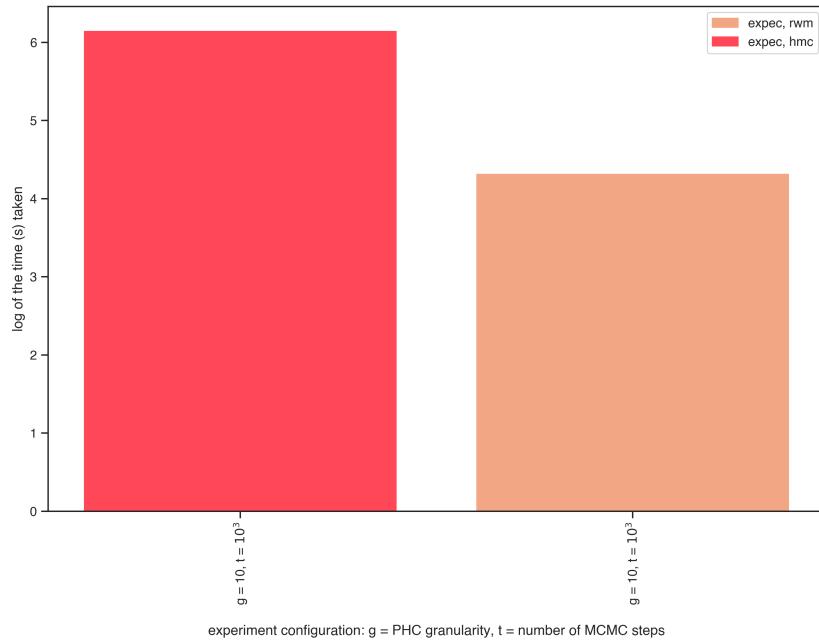


Figure A.17: Runtime of the Configurations on the $d_6 \times v_3$ Election

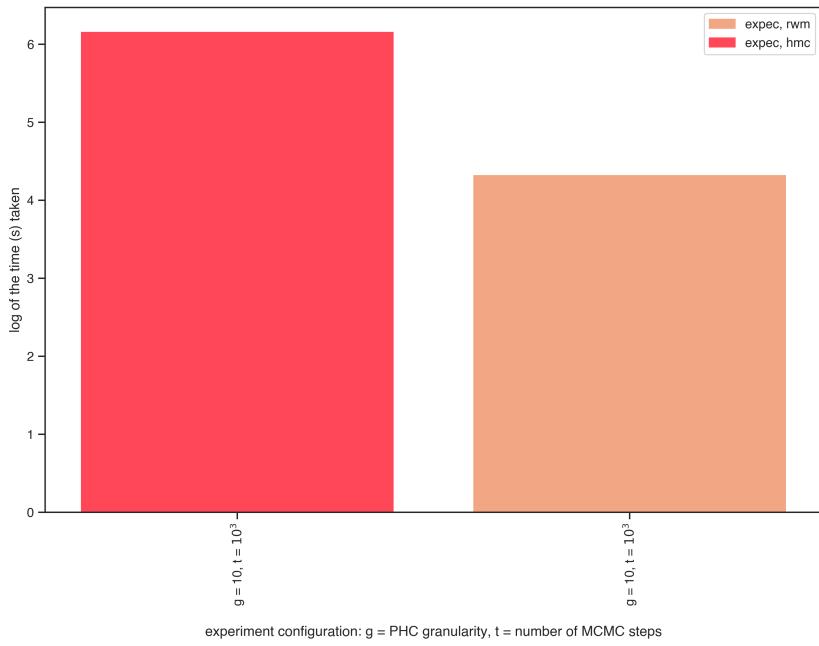


Figure A.18: Runtime of the Configurations on the $d_6 \times v_4$ Election

A.I.4 MEAN PHC MSE OF 3×2 CASES

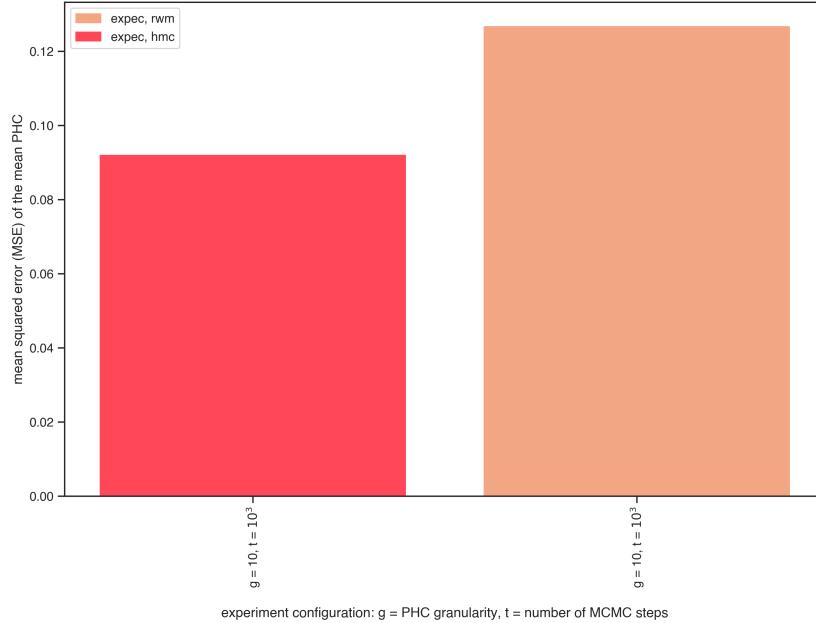


Figure A.19: MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_3$ Election

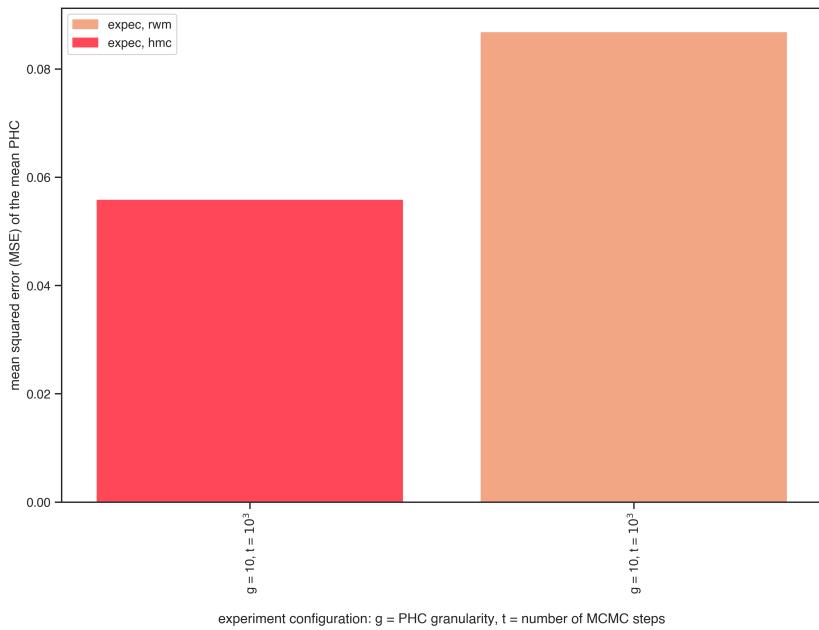


Figure A.20: MSE of the Mean PHC of the Configurations' DVM on the $d_4 \times v_4$ Election

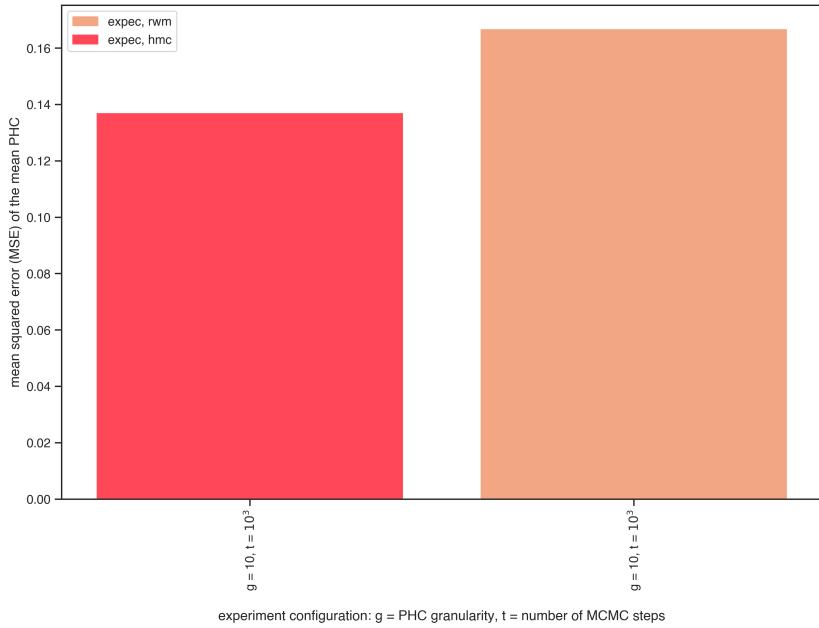


Figure A.21: MSE of the Mean PHC of the Configurations' DVM on the $d_5 \times v_3$ Election

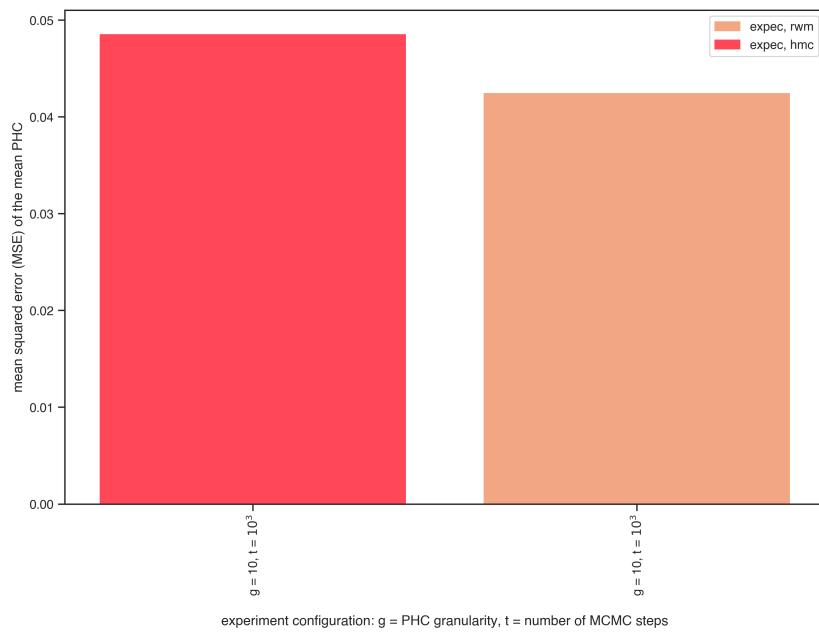


Figure A.22: MSE of the Mean PHC of the Configurations' DVM on the $d_5 \times v_4$ Election

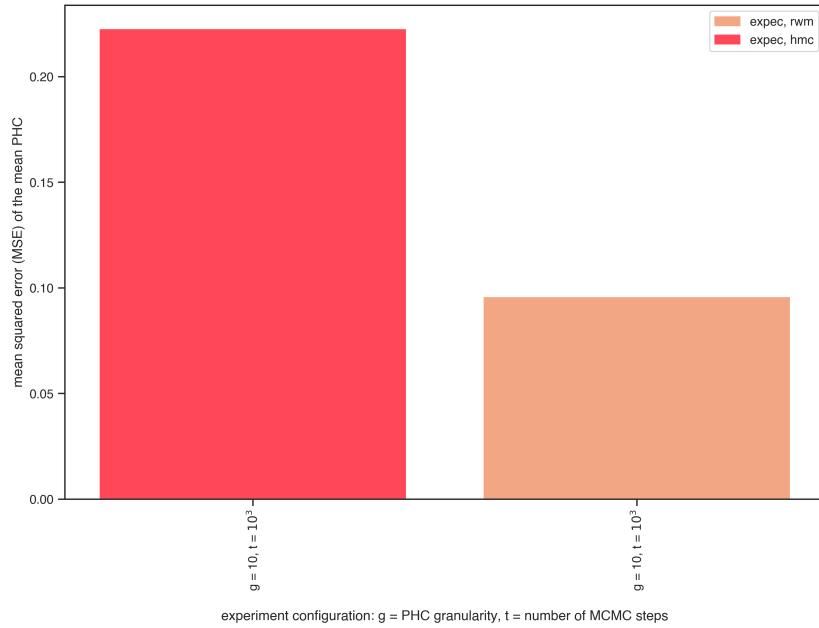


Figure A.23: MSE of the Mean PHC of the Configurations' DVM on the $d_6 \times v_3$ Election

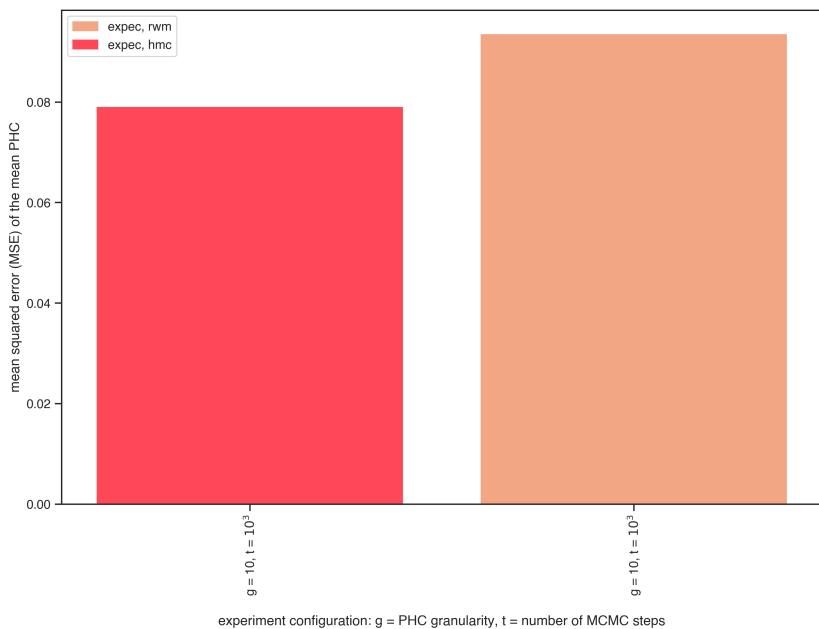


Figure A.24: MSE of the Mean PHC of the Configurations' DVM on the $d_6 \times v_4$ Election

References

- [1] Holder v. Hall, 512 U.S. 874, 904 n.13 (1994).
- [2] League of United Am. Citizens v. Clements, 999 F.2d 831, 860 (5th Cir. 1993) (en banc).
- [3] League of United Latin American Citizens v. Perry, 126 S. Ct. 2594 (2006).
- [4] Lewis v. Alamance County, 99 F.3d 600, 604 n.3 (4th Cir. 1996).
- [5] Thornburg v. Gingles. Oyez.
- [6] Voting Rights Act (1965). *Our Documents - Voting Rights Act (1965)*.
- [7] Introduction to federal voting rights laws. *The United States Department of Justice*, Aug 2015.
- [8] Is chicago mayor candidate lori lightfoot winning the latino vote?, Mar 2019.
- [9] C. Andrieu and J. Thoms. A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373, 2008.
- [10] M. Betancourt, S. Byrne, S. Livingstone, and M. Girolami. The geometric foundations of hamiltonian monte carlo. *Bernoulli*, 23(4A):2257–2298, 2017.
- [11] R. Christian and G. Casella. A short history of markov chain monte carlo: Subjective recollections from incomplete data. *arXiv.org*, 26(1), 2012.
- [12] W. Commons. File:traviscountydistricts.png — wikimedia commons, the free media repository, 2014. [Online; accessed 15-November-2019].
- [13] N. Deruelle and J.-P. Uzan. Hamiltonian mechanics. In *Relativity in Modern Physics*. Oxford University Press, 2018.
- [14] K. E. Ferree. Iterative approaches to r x c ecological inference problems: Where they can go wrong and one quick fix. 12(2), 2004.

- [15] A. Gelman, D. Park, S. Ansolabehere, P. Price, and L. Minnite. Models, assumptions and model checking in ecological regressions. *Journal Of The Royal Statistical Society Series A-Statistics In Society*, 164:101–118, 2001.
- [16] D. J. Greiner. Ecological inference in voting rights act disputes: Where are we now, and where do we want to be? *Jurimetrics*, 47(2):115–167, 2007.
- [17] G. King. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press, Princeton, New Jersey, 1997.
- [18] G. King. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press, Princeton, 2013.
- [19] G. King, O. Rosen, and M. A. Tanner. Binomial-beta hierarchical models for ecological inference. *Sociological Methods & Research*, 28(1):61–90, 1999.
- [20] J. Lao, C. Suter, I. Langmore, C. Chimisov, A. Saxena, P. Sountsov, D. Moore, R. Saurous, M. Hoffman, and J. Dillon. tfp.mcmc: Modern markov chain monte carlo tools built for modern hardware. *arXiv.org*, 2020.
- [21] F. F. Liang. *Advanced Markov chain Monte Carlo methods : learning from past samples*. Wiley series in computational statistics. Wiley, Chichester, England, 2010.
- [22] Metric Geometry and Gerrymandering Group (MGGG). Study of reform proposals for chicago city council. 2019.
- [23] R. Neal. Mcmc using hamiltonian dynamics. *arXiv.org*, 2012.
- [24] G. O. Roberts and J. S. Rosenthal. General state space markov chains and mcmc algorithms. *Probab. Surveys* 1, 1(20-71):20–71, 2004.
- [25] D. B. Rubin. *Multiple imputation for nonresponse in surveys*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, New York ;, 1987.
- [26] K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F. B. Viegas, and M. Wattenberg. Visualizing dataflow graphs of deep learning models in tensorflow. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):1–12, 2018.

A handwritten signature in black ink, appearing to read "Hakeem Olakunle Isa Angulu".

Hakeem Olakunle Isa Angulu