Predicting the 2004 Presidential Election

by

Yiqing Li

M.Sc., Huazhong Normal University, 1989

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY

March 2004

APPROVAL

Name: Yiqing Li

Degree: Master of Science

Title of project: Predicting the 2004 Presidential Election

Examining Committee: Dr. Boxin Tang
Chair

Dr. Tim Swartz
Senior Supervisor
Simon Fraser University

Dr. K. Larry Weldon
Simon Fraser University

Dr. Anil Hira
External Examiner
Simon Fraser University

Date Approved: March 1, 2004

PARTIAL COPYRIGHT LICENCE

Title of Thesis/Project/Extended Essay:

Author:

(Signature)

YIQING LI

(Name)

Jan. 30, 2004

(Date)

Abstract

This project concerns the prediction of the 2004 presidential election in the United States of America. The inferential approach is Bayesian where the primary objective is to obtain the posterior probability that the Republican party is elected. As a by-product, the posterior distribution of the number of electoral votes obtained by the Republican party is obtained. The data used in the Bayesian analysis is sample survey data from each of the 50 states and the District of Columbia, and the estimates are obtained via simulation from the posterior distribution. In addition, a sensitivity analysis is performed where $Q(100\%)$ of the Republican voters are assumed to change their vote to the Democratic party. This is done for various choices of Q.

Acknowledgements

Firstly, thanks to my senior supervisor, Dr. Tim Swartz, for his guidance, many helpful suggestions, continued support, the ongoing understanding, encouragement and motivation that has made this project possible. Secondly, thanks to all my friends in the Department of Statistics and Actuarial Science for the fun time and helpful conversations. Finally, thanks to my family for their constant support and patience throughout my education.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This project concerns the prediction of the 2004 presidential election in the United States of America (USA).

The project is motivated by commonly reported opinion polls as presented in Table 1.1. In Table 1.1 we observe that the opinion poll is based on a national sample of American voters. Although this provides some indication of the mood of the American public, it ignores the complex process (i.e. the electoral college system) used to determine the president of the USA. For example, in 2000, the Democratic candidate, Al Gore, received 49% of the popular vote, yet the presidency was awarded to the Republican, George W. Bush, who received only 48% of the popular vote. The remainder of the popular vote (3%) was cast for fringe parties (e.g. the Green Party).

Sometimes, opinion polls are reported on a state-wise basis as in Table 1.2. However, the question remains as to how to properly amalgamate the state polls keeping the electoral college process in mind. Sometimes, the electoral college process is properly considered and a projection is made. For example, it might be stated that the Republican, George W. Bush is favoured to retain his presidency (Coleman, Cantor and Neale 2000).

What we have failed to see in the literature is a probability assessment of the winner of the presidency that properly accounts for the electoral college process. For example, it would be interesting to know, for example, whether George W. Bush has a 60% chance or an 80% chance of maintaining his presidency for a second term.

This project estimates probabilities such as these based on a snapshot in time. The data used in the estimation process are the results of state sample surveys. One of the highlights of the approach is that the estimates can be readily updated whenever new surveys become available. In addition, a sensitivity analysis is performed based on hypothetical shifts in public opinion for the candidates. This may be a practical tool for political strategists who need to determine what sort of changes are required in order for their candidate to win.

In Chapter 2, we describe the electoral college system used to select the president of the United States of America. In Chapter 3, a model is proposed that uses data from state opinion polls to draw inferences on the presidential outcome. The approach is Bayesian and a simulation algorithm is proposed which allows for the estimation of posterior probabilities corresponding to various events. An advantage of a Bayesian approach is that it yields direct probability assessments of the inferential problems of interest. The sensitivity analysis is carried out in Chapter 4. The sensitivity analysis reveals key states in the presidential election. A short discussion is provided in Chapter 5.

Table 1.1: CNN/USA Today/Gallup Poll, December 5-7, 2003 based on 957 registered voters nationwide.

"If George W. Bush runs for reelection in 2004, in general, are you more likely to vote for Bush or for the Democratic Party's candidate for president?"

Bush	Democrat	Other	Not Sure
48%	41%	4%	7%

Table 1.2: Capital Survey Research Center/Alabama Education Association Poll, November 17-20, 24-25 and December 8-9, 2003 based on 735 likely voters in Alabama.

"If the general election for president were today, for whom would you vote?"

Bush	Dean	Not Sure
62%	23%	15%

Chapter 2

The Electoral College Process

The majority of the discussion appearing in Chapter 2 has been taken directly from Chapter 5 of Coleman, Cantor and Neale (2000).

2.1 The Electoral College

Every four years, the president of the United States is elected. When voters go to the polls on election day in November, they actually cast their votes for a slate of electors, who are entrusted by the Constitution with the election of the President and the Vice President. The electors are known collectively as the electoral college.

The electoral college currently includes 538 members: one for each Senator, one for each member of the House of Representatives, and three for the District of Columbia. It has no continuing existence or function apart from that entrusted to it. Each state has a number of electoral votes equal to the combined numerical total of its Senate and House delegation. Since the size of state delegations in the House of Representatives may change after the reapportionment mandated by the decennial census, the size of state representation in the electoral college has similarly fluctuated. The most recent House reapportionment and reallocation of electoral votes followed the 2000 census; the effect is that the electoral college allotment is slightly different in the election of 2004 compared to the election of 2000. Current electoral vote allocations are listed in Table 2.1 and have been obtained from Leip (2001b).

Table 2.1: 2004 electoral votes m_i by state including the District of Columbia.

State	m_i	State	m_i	State	m_i
Alabama	9	Louisiana	9	Ohio	20
Alaska	3	Maine	4	Oklahoma	7
Arizona	10	Maryland	10	Oregon	7
Arkansas	6	Massachusetts	12	Pennsylvania	21
California	55	Michigan	17	Rhode Island	4
Colorado	9	Minnesota	10	South Carolina	8
Connecticut	7	Mississippi	6	South Dakota	3
Delaware	3	Missouri	11	Tennessee	11
Florida	27	Montana	3	Texas	34
Georgia	15	Nebraska	5	Utah	5
Hawaii	4	Nevada	5	Vermont	3
Idaho	4	New Hampshire	4	Virginia	13
Illinois	21	New Jersey	15	Washington	11
Indiana	11	New Mexico	5	West Virginia	5
Iowa	7	New York	31	Wisconsin	10
Kansas	6	North Carolina	15	Wyoming	3
Kentucky	8	North Dakota	3	District of Columbia	3
				Total:	538

It is apparent that the populous states (e.g. California, New York, Texas) have many more electors than the smaller states (e.g. Delaware, Wyoming).

The Constitution does not specify procedures for the nomination of candidates for the office of presidential elector. The states have adopted various methods of nomination for elector candidates, of which the two most popular are by state party convention, used in 36 states, and by state party committee, used in 10 states. In practice, elector candidates tend to be prominent state and local officeholders, party activists, and other citizens associated with the party to which they are nominated. A key point is that there is a "ticket" of electors for each party (e.g. Republican, Democrat) with each elector pledged to a particular party. The number of electors in the ticket is equal to the number of electoral votes for the state. The method of selecting electors and of awarding electoral votes has been left to the states. However, in 48 states and the District of Columbia, all electoral votes are awarded to the party that receives a plurality of popular votes in the state. This practice is variously known as the general ticket or winner-take-all system. The general ticket system usually tends to exaggerate the winning candidates' margin of victory, as compared with the share of popular votes received. For instance, in 1996, Bill Clinton and Al Gore won 49.2% of the popular vote, as compared with 40.7% by Bob Dole and Jack Kemp. The Democrats' electoral vote margin of 379 to 159 was a much higher 70.4% of the total, due to the fact that the Democratic ticket received a plurality vote in 32 states and the District of Columbia.

2.2 Winning the Presidency

To win the presidency, the candidate must receive a majority of electoral votes (currently 270 of the 538 total). Once the voters have chosen the members of the electoral college, the electors meet to ratify the voters' choices for President and Vice President.

In the rare event that an electoral college majority is not attained (e.g. 239 electoral votes are cast for both Republicans and Democrats), then it is left to the House of Representatives to elect the President.

The founding fathers intended that individual electors be free agents, voting for the candidates they thought most fit to hold office. In practice, however, electors are not expected to exercise their own judgment, but, rather, to ratify the people's choice by voting for the candidates winning the most popular votes in their state. Despite this understanding, there is no constitutional provision requiring electors to vote as they have pledged. Over the years, a number of electors have voted against the voters' instructions, known as the phenomenon of the unfaithful, or faithless, elector. Although a number of states have laws which seek to bind the electors to the popular vote winners, the preponderance of opinion among constitutional scholars holds that electors remain free agents. Moreover, all of the seven votes of the faithless electors between 1948 and 1988 were recorded as cast. The most recent occurrence was in 1988, when a West Virginia Democratic elector voted for the Republican Lloyd Bentsen for President and Michael Dukakis for Vice President.

Chapter 3

A Bayesian Model for Prediction

In this chapter, we propose a Bayesian model for use in the prediction of the 2004 presidential election in the USA.

We recall from Chapter 2 that the electoral college system determines the presidency. We therefore define p_i as the proportion of all voters in the i-th state at the time of this state's survey who would vote Republican, $i = 1, \ldots, 51$. Note that p_i is an unknown parameter and that a Bayesian approach permits inference on p_i where p_i is regarded as random. Note also that $p_1, \ldots, p_{50}$ are ordered alphabetically according to the state name and that p_{51} is the unknown proportion corresponding to the District of Columbia.

We introduce the generic notation $[A]$ to denote either the density or probability mass function of a random variable A. Similarly, we let $[A|B]$ denote either the density or probability mass function of the random variable A conditional on the random variable B. Using this notation and Bayes theorem, we express the joint posterior distribution of $p_1, \ldots, p_{51}$ given survey data $\underline{y}_1, \ldots, \underline{y}_{51}$ via

$$\left[p_1, \ldots, p_{51} | \underline{y}_1, \ldots, \underline{y}_{51}\right] \propto \left[\underline{y}_1, \ldots, \underline{y}_{51} | p_1, \ldots, p_{51}\right] [p_1, \ldots, p_{51}] \tag{3.1}$$

Using a conditional independence assumption

$$\left[\underline{y}_1, \ldots, \underline{y}_{51} | p_1, \ldots, p_{51}\right] = \prod_{i=1}^{51} \left[\underline{y}_i | p_1, \ldots, p_{51}\right]$$

that implies that state survey results are independent given state population proportions and imposing a vague (flat) prior on the proportions, (3.1) reduces to

$$\left[p_1,\ldots,p_{51}|\underline{y}_1,\ldots,\underline{y}_{51}\right] \propto \prod_{i=1}^{51}\left[\underline{y}_i|p_i\right]$$

which implies

$$\left[p_i|\underline{y}_1,\ldots,\underline{y}_{51}\right] = \left[p_i|\underline{y}_i\right] \propto \left[\underline{y}_i|p_i\right], i = 1,...,51. \tag{3.2}$$

Now let us jump ahead for a moment and assume that the distributions in (3.2) are tractable in the sense that random variate generation is straightforward. We then define X_i as the number of electoral votes for the Republican party corresponding to the i-th state, $i = 1,\ldots,51$. Assuming that the proportion p_i of Republican voters at the time of the survey is the same as the proportion of Republican voters at the time of the general election (an assumption which we relax in Chapter 4), the rules of the electoral college system give

$$X_i = \begin{cases} m_i & \text{if } p_i > 0.5 \\ 0 & \text{if } p_i \leq 0.5 \end{cases} \tag{3.3}$$

where m_i is the number of electoral college votes assigned to state i as shown in Table 2.1 $i = 1,\ldots,51$.

With 538 total electoral college votes (see Table 2.1), the Republicans win the 2004 presidential election outright if

$$T = \sum_{i=1}^{51} X_i \geq 270.$$

Refer again to Chapter 2 for a discussion of the case where $T = 239$.

The theory outlined above suggests an algorithm for estimating the posterior probability that the Republicans are elected $\left(\text{i.e. } P(T \geq 270|\underline{y}_1,...,\underline{y}_{51})\right)$:

- Step 0 : Set Counter $= 0$
- Step 1 : Repeat for $j = 1,\ldots,M$

- For $i = 1, \ldots, 51$, generate $p_i \sim \left[p_i | \underline{y}_1, ..., \underline{y}_{51}\right]$ as in (3.2)
- For $i = 1, \ldots, 51$, set X_i as in (3.3)
- If $T = \sum_{i=1}^{51} X_i \geq 270$, Counter = Counter +1

- Step 2: $\widehat{P}(T \geq 270 | \underline{y}_1, ..., \underline{y}_{51}) =$ Counter $/M$

For the simulations appearing later in the project, we choose M=1000. Clearly the algorithm can also be used for estimating the posterior distribution of T (i.e. the number of electoral votes obtained by the Republican party). This is simply done by saving the generated T-values $T_1, \ldots, T_M$ and forming a histogram.

Although the proposed approach and algorithm are simple, there are three complicating details that need to be worked out and we discuss these in sections 3.1, 3.2 and 3.3 respectively.

3.1 Maine and Nebraska

We quote directly from Leip (2001a):

> The candidate with the highest popular vote tally receives all of the states electoral votes with the exception of electoral votes from Maine and Nebraska, which use the District Method: Two Electors At Large and one Electors for each Congressional District are pledged to each Presidential / Vice Presidential ticket (the Electors' names may or may not appear on the ballot). The At-Large Electors pledged to the ticket having received the plurality of votes state-wide are chosen. The Elector pledged to the ticket having received the plurality of votes within each Congressional District is chosen.

Since we are unsure as to the precise workings of the Congressional Districts, we have decided to treat these two states in the same manner as the remaining 49 "states." As the number of electoral votes for Maine and Nebraska are small (4 and 5 respectively), this should have a minor impact on our inferences.

3.2 Sampling from the Distributions in (3.2)

In order to sample from the distributions in (3.2), it is first necessary to look at the form of the state surveys. The surveys produce count data and take various forms. Some simplified examples are shown below:

(a) Would you vote for the Republican or the Democrat?

(b) Would you vote for George Bush or the Democratic if the Democratic nominee were Howard Dean?

(c) Would you vote for the Republican, the Democrat or someone else?

In all of the surveys that are used in this project, the support for parties other than the Republican Party and the Democratic Party, when reported, has been negligible. In most cases, the surveys group undecided voters with those who support the fringe parties. Our data $\underline{y}_i$, $i = 1, ..., 51$ has been reduced to the common form:

$$\underline{y}_i = (y_{i1}, y_{i2}, y_{i3}) \sim \text{Multinomial}\,(n_i, q_{i1}, q_{i2}, q_{i3}) \tag{3.4}$$

where

- y_{i1} = number of survey votes for the Republican
- y_{i2} =number of survey votes for the Democrat
- y_{i3} = number of survey votes for other party or unsure
- n_i = survey sample size

Because of the minimal support for the fringe parties, we treat y_{i3} as the number of undecided voters in the survey. To maximize the uncertainty due to the undecided voters, we modify our data in (3.4) by first generating

$$w_i \sim \text{Binomial}\,(y_{i3}, 1/2) \tag{3.5}$$

yielding (see Appendix)

$$y_{i1} + w_i \sim \text{Binomial}\,(n_i, p_i)$$

where $p_i = 1/2 + q_{i1}/2 - q_{i2}/2$. This has the effect of dispersing the undecided vote to the two parties.

Viewing the Binomial density as function of the parameter p_i, it follows that the posterior distributions in (3.2) are given by

$$\left[p_i|\underline{y}_i\right] \sim \text{Beta}\,(y_{i1} + w_i + 1, n_i - y_{i1} - w_i + 1) \tag{3.6}$$

for $i = 1, ..., 51$. It therefore simple to generate the p_i variates in the algorithm; first generate w_i according to (3.5) and then generate p_i according to (3.6).

3.3 States without Survey Data

When our survey data were first collected (January 10, 2003), there were 12 states for which no survey data existed (see Table 3.1). Note that these are small states (politically) with few electors.

This does not prove to be a problem for our methodology as the default Bayesian approach treats the prior as the posterior. To be more precise, if the i-th state has no survey data then the generation steps (3.5) and (3.6) reduce to

$$\left[p_i|\underline{y}_i\right] \sim \text{Beta}(1,1) \sim \text{Uniform}(0,1)$$

where $y_{i1} = y_{i2} = y_{i3} = 0$ and $w_i = 0$.

Perhaps a preferable approach is to group these states with comparable states. By comparable we mean states that are geographically close (if possible) and politically close. See Table 3.1 for our proposed matchings. The information in Table 3.1 was obtained from Leip (2001c). Using this approach, we combine the number of electoral votes in the matching states and use survey data for the member in the pair for whom survey data exists. The intuition is that the paired states will have the same election results (i.e. both paired states will either have a Republican or a Democrat as the winner). Note that the District of Columbia was very heavily Democratic in the 2000 election and there are really no comparable states politically. We will use the matching procedure in all of our subsequent analyses.

Table 3.1: The 12 "states" for which survey data did not exist on January 10, 2004. The states are matched with states that are comparable geographically and politically.

State	Electors m_i	% Republican in 2000	Matching State	% Republican in 2000
Alaska	3	58.62%	Texas	59.30%
Delaware	3	41.90%	Maryland	40.18%
Hawaii	4	37.46%	California	41.65%
Idaho	4	67.17%	Montana	58.44%
Kansas	6	58.04%	Oklahoma	60.31%
Mississippi	6	57.62%	Alabama	56.47%
North Dakota	3	60.66%	Nebraska	62.25%
South Dakota	3	60.30%	Nebraska	62.25%
Utah	5	66.83%	Montana	58.44%
West Virginia	5	51.92%	Virginia	52.47%
Wyoming	3	67.76%	Montana	58.44%
District of Columbia	3	8.95%	Maryland	40.18 %

3.4 Results (January 10, 2004)

The data are presented in Table 3.2. The data were obtained from the Polling Report website (www.politics1.com) collected on January 10, 2004. This website provides extensive polling results for many topics of interest in the USA. To access the subscriber pages, a user must pay a small fee.

The Bayesian analysis was carried out using the simulation algorithm and the estimated posterior probability of the election of a Republican president is 1.00. Of course, we remind the reader that this calculation is based on the unrealistic assumption that the American voting public does not change their opinion from the time of the survey to election time. A sensitivity analysis is performed in Chapter 4 which addresses this issue. In Figure 3.1 we provide a histogram depicting the posterior distribution of T, the total number of the Republican electoral college votes. We observe that if sentiment at election time remains the same as on January 10, 2004, then with high probability the Republicans will obtain somewhere between 314 and

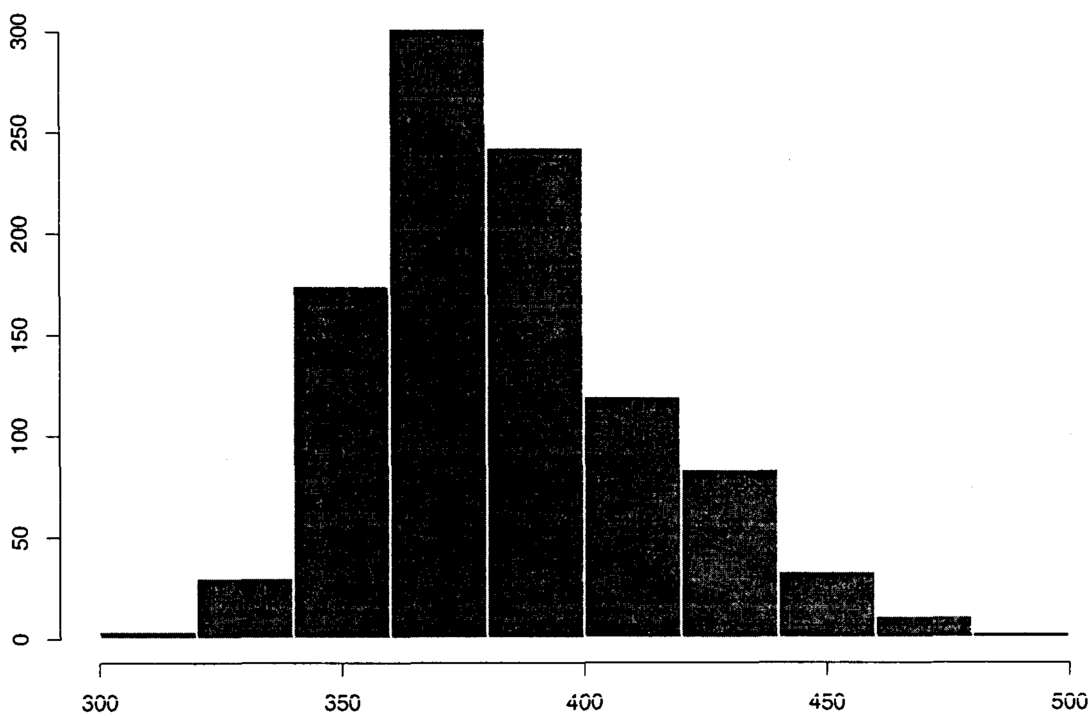

Figure 3.1: Histogram of the total number of Republican electoral votes T based on the survey data from January 10, 2004.

489 electoral votes. The mean number of Republican electoral votes is 385.

3.5 Results (February 2, 2004)

The data were collected again on February 2, 2004 and the survey results are shown in Table 3.3. There are 21 new surveys when compared to January 10, 2004. The Bayesian analysis was performed again and the estimated posterior probability of the election of a Republican remains 1.00. In Figure 3.2, we present the corresponding posterior histogram of T. This time we observe that if sentiment at election time remains the same as on February 2, 2004, then with high probability the Republicans

Table 3.2: State sample survey data obtained on January 10, 2004.

State	Survey sample size n_i	Republican votes y_{i1}	Democrat votes y_{i2}	Undecided y_{i3}
Alabama	735	456	169	110
Arizona	409	205	143	61
Arkansas	900	324	171	405
California	500	210	235	55
Colorado	900	297	225	378
Connecticut	508	229	173	106
Florida	800	392	328	80
Georgia	500	245	205	50
Illinois	900	342	324	234
Indiana	803	618	112	73
Iowa	600	288	258	54
Kentucky	801	521	264	16
Louisiana	700	462	196	42
Maine	600	216	270	114
Maryland	806	330	387	89
Massachusetts	401	136	233	32
Michigan	600	276	216	108
Minnesota	900	252	225	423
Missouri	600	294	234	72
Montana	625	344	206	75
Nebraska	604	477	121	6
Nevada	625	319	275	31
New Hampshire	600	306	204	90
New Jersey	604	272	272	60
New Mexico	400	208	136	56
New York	706	289	325	92
North Carolina	600	324	240	36
Ohio	593	338	225	30
Oklahoma	500	300	125	75
Oregon	400	164	136	100
Pennsylvania	1092	535	470	87
Rhode Island	367	136	121	110
South Carolina	628	327	226	75
Tennessee	400	208	148	44
Texas	994	616	308	70
Vermont	400	152	200	48
Virginia	625	331	200	94
Washington	500	220	180	100
Wisconsin	510	235	240	35

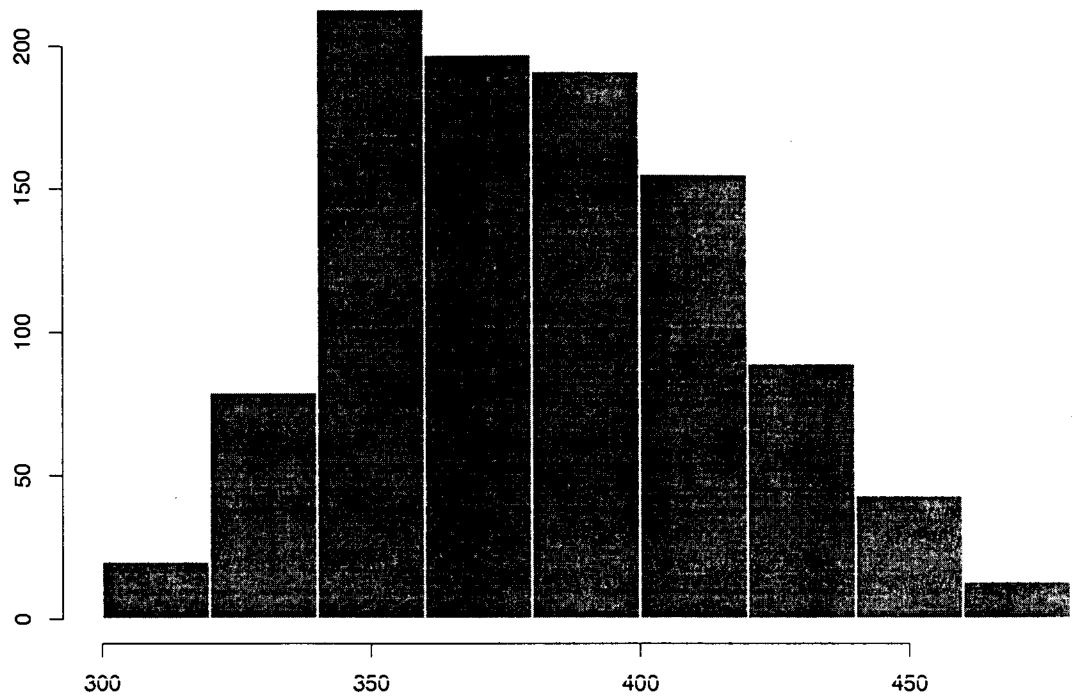

Figure 3.2: Histogram of the total number of Republican electoral votes T based on the survey data from February 2, 2004.

will obtain somewhere between 300 and 472 electoral votes. The mean number of Republican electoral votes is 382.

Table 3.3: State sample survey data obtained on February 2, 2004. Asterisks marked beside states indicate states for which survey data did not exist on January 10, 2004.

State	Survey sample size n_i	Republican votes y_{i1}	Democrat votes y_{i2}	Undecided y_{i3}
Alabama	735	456	169	110
Arizona	641	321	244	76
Arkansas	900	324	171	405
California	929	427	437	65
Colorado	400	184	144	72
Connecticut	1600	736	752	112
Florida	800	344	296	160
Georgia	500	245	205	50
Illinois	700	280	336	84
Indiana	803	618	112	73
Iowa	600	288	258	54
Kentucky	801	521	264	16
Louisiana	607	304	225	78
Maine	400	176	124	100
Maryland	1200	490	612	96
Massachusetts	405	126	223	56
Michigan	400	204	140	56
Minnesota	625	256	269	100
Missouri	804	362	322	120
Montana	625	344	206	75
Nebraska	604	477	121	6
Nevada	625	319	275	31
New Hampshire	601	325	240	36
New Jersey	904	371	344	189
New Mexico	400	160	172	68
New York	617	210	222	185
North Carolina	600	318	240	42
Ohio	593	338	225	30
Oklahoma	500	300	125	75
Oregon	400	164	136	100
Pennsylvania	1092	535	470	87
Rhode Island	367	136	121	110
South Carolina	807	363	347	97
South Dakota*	400	196	88	116
Tennessee	400	208	148	44
Texas	900	459	153	288
Vermont	400	152	200	48
Virginia	625	331	200	94
Washington	500	220	180	100
Wisconsin	510	235	240	35

Chapter 4

Sensitivity Analysis

In this chapter, we use our most recent set of data, that is the sample survey data that was collected on February 2, 2004 and is summarized in Table 3.3.

We saw in Chapter 3 that if public sentiment does not change between the survey dates and the time of the general election, then with certainty, George W. Bush will be elected for a second term of office. We are interested in the degree to which public sentiment needs to shift in order for the election results to change. We see this sort of information as a potentially valuable tool for campaign strategists. We emphasize again that whenever new survey results become available, the analyses can be easily updated.

We therefore consider a shift of $Q(100\%)$ of Republican voters to the Democratic side where we choose $Q = 0.00, 0.01, \ldots, 0.10$. Note that the choice $Q = 0.00$ refers to the analysis of Section 3.5, $Q = 0.01$ refers to a 1% shift and $Q = 0.10$ refers to a 10% shift.

Our algorithm changes only slightly for general Q and appears below.

- Step 0 : Set Counter $= 0$
- Step 1 : Repeat for $j = 1, \ldots, M$
  - For $i = 1, \ldots, 51$, generate $w_i \sim \text{Binomial}(y_{i3}, 1/2)$
  - For $i = 1, \ldots, 51$, generate $p_i \sim \text{Beta}(y_{i1} + w_i + 1, n_i - y_{i1} - w_i + 1)$

- For $i = 1, \ldots, 51$, set $X_i = 0$
- For $i = 1, \ldots, 51$, set $X_i = m_i$if $(1 - Q)p_i > 0.5$
- If $T = \sum_{i=1}^{51} X_i \geq 270$, Counter $=$ Counter $+1$

- Step 2: $\widehat{P}(T \geq 270 | \underline{y}_1, \ldots, \underline{y}_{51}) =$ Counter$/M$

Again, posterior distributions for T under the hypothesized shift Q are simply obtained by forming a histogram from the generated T-variates.

In Table 4.1 we summarize the results of our sensitivity analysis. We observe a reversal in the election outcome at the value $Q = 0.06$. We also observe that the reversal is dramatic in the narrow range $Q = 0.05$ to $Q = 0.07$ as the probabilities change greatly from 0.80 to 0.25.

In Table 4.2 we break down the case $Q = 0.06$ presenting the posterior means of $(1 - Q)p_i$ for each state $i = 1, \ldots, 51$. We also present the cumulative total of Republican electoral votes where the states are ordered according to $(1 - Q)p_i$. We observe that the dramatic change occurs in the key states of Florida and Pennsylvania which have 27 and 21 electoral votes respectively. These states have the required number of votes to tip the presidency from Republican to Democrat.

Table 4.1: Posterior probabilities of a Republican victory based on a swing of Q (100%) of Republican voters to Democrats.

Q	$\widehat{P}(T \geq 270 \mid \underline{y}_1, \ldots, \underline{y}_{51})$
0.00	1.00
0.01	1.00
0.02	1.00
0.03	0.98
0.04	0.93
0.05	0.80
0.06	0.50
0.07	0.25
0.08	0.09
0.09	0.01
0.10	0.01

Table 4.2: Posterior means of Republican support $(1-Q)p_i$ reflecting the swing of $Q(100\%)$ Republicans to Democrats where $Q = 0.06$. We list the states in decreasing order of $(1-Q)p_i$. The number of cumulative Republican electoral votes is also provided.

State (m_i)	Cumulative Votes	$(1-Q)p_i$	State (m_i)	Cumulative Votes	$(1-Q)p_i$
Indiana (11)	11	0.77	Missouri (11)	302	0.49
Nebraska (5)	16	0.75	Rhode Island (4)	306	0.49
Alabama (9)	25	0.65	New Jersey (15)	321	0.48
Oklahoma (7)	32	0.63	Utah (5)	326	0.48
Texas (34)	66	0.63	South Carolina (8)	334	0.48
Kentucky (8)	74	0.62	Mississippi (6)	340	0.48
South Dakota (3)	77	0.60	Wyoming (3)	343	0.48
Montana (3)	80	0.57	Hawaii (4)	347	0.47
Virginia (13)	93	0.57	Kansas (6)	353	0.47
Ohio (20)	113	0.56	Alaska (3)	356	0.47
Arkansas (6)	119	0.55	West Virginia (5)	361	0.47
Michigan (17)	136	0.54	Idaho (4)	365	0.47
Tennessee (11)	147	0.54	D.C. (3)	368	0.47
New Hampshire (4)	151	0.54	Wisconsin (10)	378	0.47
Maine (4)	155	0.53	Connecticut (7)	385	0.47
Louisiana (9)	164	0.53	California (55)	440	0.47
North Carolina (15)	179	0.53	New York (31)	471	0.46
Arizona (10)	189	0.53	North Dakota (3)	474	0.46
Colorado (9)	198	0.52	Minnesota (10)	484	0.46
Georgia (15)	213	0.51	Delaware (3)	487	0.46
Washington (11)	224	0.51	New Mexico (5)	492	0.46
Oregon (7)	231	0.50	Illinois (21)	513	0.43
Nevada (5)	236	0.50	Maryland (10)	523	0.42
Florida (27)	263	0.50	Vermont (3)	526	0.41
Pennsylvania (21)	284	0.50	Massachusetts (12)	538	0.36
Iowa (7)	291	0.49			

Chapter 5

Concluding Remarks

This project develops a Bayesian approach based on simulation for obtaining probability assessments of attaining the presidency. There is very little of a subjective nature to the analysis as it is based entirely on survey data which has historically been a reliable source of information.

One of the nice features of the approach is that assessments can be readily updated as new surveys are obtained. An important assumption in the approach is that the voting tendencies at election time will be the same as when the surveys are obtained. To address this obvious deficiency, a sensitivity analysis is carried out where fixed percentages of Republicans change their allegiance to the Democratic side.

Although our sensitivity analysis considered the swing of Republicans to Democrats on a national level, it would be a simple matter to allow the swings Q_i to depend differentially by state. This may be sensible when political parties campaign more heavily in different regions.

In the time since the survey data were collected, the Independent candidate Ralph Nader has entered the race for the Presidency. In light of this revelation, it may be sensible to apportion the undecided vote w_i differently. It is widely accepted that some of the left leaning vote that would normally go to the Democrats will now go to Nader.

It was also suggested by the external examiner Dr. Hira that in a close race, there is traditionally a higher voter turnout, and the higher numbers may help the Democratic cause. Perhaps we could model this phenomena to improve the estimation.

Finally, we urge readers not to take the results too seriously. At the time of the surveys, a Democrat candidate had not yet been determined. We expect significant changes in survey results after the Democratic nomination is known.

Appendix

Proposition: If $(y_1, y_2, y_3) \sim$Multinomial(n, q_1, q_2, q_3) and $w \sim$Binomial$(y_3, 1/2)$, then $y_1 + w \sim$Binomial$(n, 1/2 + q_1/2 - q_2/2) \sim$Binomial$(n, q_1 + (1 - q_1 - q_2)/2)$.

Proof: Consider the joint distribution given by

$$\begin{aligned} [w, y_1, y_2] &= [w|y_1, y_2]\,[y_1, y_2] \\ &= \binom{n-y_1-y_2}{w}\left(\tfrac{1}{2}\right)^{n-y_1-y_2} \frac{n!}{y_1!y_2!(n-y_1-y_2)!} q_1^{y_1} q_2^{y_2} \left(1 - q_1 - q_2\right)^{n-y_1-y_2} \end{aligned}$$

where $0 \le y_1, 0 \le y_2,\ \ y_1 + y_2 \le n,\ \ 0 \le w \le n - y_1 - y_2$. Let $t = y_1 + w$. Then the joint distribution is given by

$$[t, y_1, y_2] = \binom{n-y_1-y_2}{t-y_1}\left(\tfrac{1}{2}\right)^{n-y_1-y_2} \frac{n!}{y_1!y_2!(n-y_1-y_2)!} q_1^{y_1} q_2^{y_2} \left(1 - q_1 - q_2\right)^{n-y_1-y_2}$$

where $0 \le y_2,\ \ 0 \le y_1 \le t,\ \ t + y_2 \le n$.

The marginal distribution is then given by

$$
\begin{aligned}
[t] &= \sum_{y_1=0}^{t}\sum_{y_2=0}^{n-t}[t, y_1, y_2] \\
&= \sum_{y_1=0}^{t} \frac{n!}{y_1!(t-y_1)!}\left(\tfrac{1}{2}\right)^{n-y_1} q_1^{y_1}(1-q_1-q_2)^{n-y_1}\sum_{y_2=0}^{n-t}\frac{1}{y_2!(n-y_2-t)!}\left(\tfrac{1}{2}\right)^{-y_2} q_2^{y_2}(1-q_1-q_2)^{-y_2} \\
&= \sum_{y_1=0}^{t} \frac{n!}{y_1!(t-y_1)!}\left(\tfrac{1}{2}\right)^{n-y_1} q_1^{y_1}(1-q_1-q_2)^{n-y_1}\frac{1}{(n-t)!}\frac{1}{(1-q_1-q_2)^{n-t}} \\
&\quad \cdot \sum_{y_2=0}^{n-t}\frac{(n-t)!}{y_2!(n-y_2-t)!}(2q_2)^{y_2}(1-q_1-q_2)^{n-t-y_2} \\
&= \sum_{y_1=0}^{t}\frac{n!}{y_1!(t-y_1)!(n-t)!}\left(\tfrac{1}{2}\right)^{n-y_1} q_1^{y_1}(1-q_1-q_2)^{t-y_1}(2q_2+1-q_1-q_2)^{n-t} \\
&= \frac{n(n-1)...(t+1)!}{(n-t)!}\sum_{y_1=0}^{t}\binom{n}{y_1}\left(\tfrac{1}{2}\right)^{n-y_1} q_1^{y_1}(1-q_1-q_2)^{t-y_1}(1-q_1+q_2)^{n-t} \\
&= \binom{n}{t}(1-q_1+q_2)^{n-t}\left(\tfrac{1}{2}\right)^{n-t}\sum_{y_1=0}^{t}\binom{t}{y_1}\left(\tfrac{1}{2}\right)^{t-y_1} q_1^{y_1}(1-q_1-q_2)^{t-y_1} \\
&= \binom{n}{t}(1-q_1+q_2)^{n-t}\left(\tfrac{1}{2}\right)^{n-t}(q_1+1/2-q_1/2-q_2/2)^{t} \\
&= \binom{n}{t}(1/2-q_1/2+q_2/2)^{n-t}(1/2+q_1/2-q_2/2)^{t}
\end{aligned}
$$

where $0 \le t \le n$. Hence $t \sim$Binomial$(n, 1/2 + q_1/2 - q_2/2)$.

Bibliography

[1] Coleman, K.J., Cantor, J.E., Neale, T.H. (2000). Presidential Elections in the United States: A Primer, www.senate.gov/reference/resources/pdf/RL30527.pdf.

[2] Leip, D. (2001a). Methods of Choosing Presidential Electors, from following website: www.uselectionatlas.org/INFORMATION/INFORMATION/methods.php.

[3] Leip, D. (2001b). Electoral College Calculator - 2004, from following website: www.uselectionatlas.org/USPRESIDENT/evcalc.php?year=2004.

[4] Leip, D. (2001c). 2000 Presidential Election Results, from following website: www.uselectionatlas.org/USPRESIDENT/.