

Methodology

There are two parts in the computation:

- Computing the distribution of electoral colleges for a candidate, given the win/loss probability of each state for this candidate.
- Using state-by-state poll data to infer the win/loss probability of each state for this candidate

Quantitative Evaluation of the Distribution of Electoral Colleges for a Candidate

We provide mathematical details to compute the probability distribution of a candidate's (forecast) electoral college votes. For concreteness, we will examine Kerry's electoral votes (with Bush's count being 538 minus Kerry's).

We define the following quantity:

$k =$ index for each state, $k = 1, 2, 3, \dots, 51$ (50 states plus D.C.)

We have ordered the state alphabetically. For example, $k = 1$ corresponds to Alabama.

$n_k =$ # of electoral votes for "state" k . For example, it equals 9 for Alabama

$q_k =$ probability that Kerry will win state k : getting more than 50% of the popular votes in that state. This is the probability that Kerry will win n_k electoral votes

$P(j, S) =$ probability that Kerry wins j electoral votes considering only the first S "states".

When $S = 51$, we have included all the states in our calculation.

$P(j, 51), j = 0, 1, 2, \dots, 538$, is the distribution of Kerry's electoral vote count.

The probability that Kerry wins the presidency is given as:

$$P(\text{Kerry wins the election}) = \sum_{j=270}^{538} P(j, 51).$$

$$P(\text{Bush wins the election}) = \sum_{j=0}^{268} P(j, 51).$$

$$P(\text{Tie}) = P(269, 51).$$

See *Tie-Breaking rule* in Appendix A at the end of this document.

We assume that Kerry's performance (win/loss) in each state is independent from the outcomes in other states. This is a plausible assumption, if we condition the state-by-state outcomes on all the current events (e.g., last minute late development on Iraq, or Bin Laden, or other development) – since we will be using the up-to-date information to compute each p_k .

The following recursion is used to iteratively compute $P(j, k)$ if we know $P(j, k - 1)$:

$$P(j, k) = q_k P(j - n_k, k - 1) + (1 - q_k) P(j, k - 1), \quad j = 0, 1, 2, 3, \dots, 538; \quad k = 1, 2, 3, 4, \dots, \dots, 51.$$

$$P(0, 0) = 1, \quad P(j, 0) = 0, \text{ for all } j; \quad P(j, k) = 0, \quad j < 0; \quad k = 0, 1, 2, 3, 4, \dots, \dots, 51.$$

This recursion implicitly “aggregates” all the relevant probabilities without having to exhaustively examine all combinations of win/loss scenarios.

Using state-by-state Polling Data to infer a Candidate’s Win/Loss Probability for this State

We will not carry the state index k in this section. The analysis here will be carried out for each state.

We do not know how each state will vote, in terms of the proportion of (popular) votes that will go to Kerry (for example). We seek to guess this proportion by articulating a probability density function. Suppose we denote this (unknown/uncertain) proportion as \mathbf{x} , \mathbf{x} can range in the interval $(0, 1)$. We will use conditional probability (or the Bayes’ rule) to update the probability density function for \mathbf{x} (denoted as $f(x)$), based on polling data for that state. If we know the (updated) density function for \mathbf{x} , we can compute the probability that Kerry will win this state by evaluating:

$$P(\text{Kerry wins this state}) = q = \int_{x=0.5}^1 f(x) dx.$$

This integral simply calculates the chance that the proportion of (popular) votes going to Kerry is above 50%.

A natural density function to use is the *Beta* distribution, which provides easy updating. A *Beta* distribution is determined by two parameters: A, B . Once specified, we can perform the integration. The parameters A and B correspond naturally to the polling data: how many respondents in the sample survey favor one or the other candidate.

$$f(\mathbf{x} | A, B) = K(A, B) \mathbf{x}^{A-1} (1 - \mathbf{x})^{B-1}, \quad 0 \leq \mathbf{x} \leq 1.$$

$K(A, B)$ is a normalizing constant so that the density integrates to one. This is a two-candidate model (ignoring any third party candidate), with A being the number of respondents in the survey favoring Kerry, and B favoring Bush. Once the parameters A and B are determined, we can integrate to obtain q .

There still remains the “aggregation” of many polls of different vintages to determine A and B . The traditional application of conditional probability using the *Beta* distribution and random sampling simply “adds” all the poll results (keeping track of all the Kerry/Bush sample poll votes). This seems problematic because the population proportion is not (and should not be) static. It is subject to change (particularly the undecided) as a result of current events. While a rigorous analysis is being investigated, we will use a heuristic approach to “aggregate” all the polling results (separately for each state), by using an exponential decay to discount distant (past) polling data – the discount is a function of the “age” of a poll, relative to the most recent poll. Consider a (time) series of polls $\{m_t\}$, where $m_t = \#$ of Kerry supporters in a poll conducted at time t . Suppose the current time is c , we compute the parameter $A(c)$ as:

$$A(c) = \sum_{t \leq c} e^{\frac{-(c-t)}{h}} m_t.$$

The constant $h > 0$ can be viewed as a “decaying” parameter – how fast (or slow) the past poll data will be discounted. If h is large, the “decay” is slow, meaning that past data will be discounted less. On the other hand, a small h discounts last poll results heavily. For a very small h , only the most recent poll will be counted. We can similarly compute the $B(c)$ parameter by keeping track of (the number of) Bush supporters in the polls. The unit of time (c and t) is measured in days as well. In our web site, we have judiciously chosen $h = 2$ days – particularly when the election is imminently approaching when many things are changing.

Data Source and How We Update our Data

We have chosen to download poll data from <http://www.race2004.net>. A program is written to automatically download poll data from this site and update our electoral vote distribution calculation.

Appendix A

What happens in case the Electoral College is tied (for 2004, this corresponds to 269 votes for each candidate)?

The Twelfth Amendment, proposed by Congress in 1803 and ratified by the states the next year (following the 1800 tie in the Electoral College), provides that if no presidential candidate has obtained a simple majority of the votes in the Electoral College, the House of Representatives elects the President according to the following procedure. Only the three presidential candidates with the highest number of votes remain in the race (unless Ralph Nader obtained at least one vote in the Electoral College, this would mean only John Kerry and George W. Bush would remain in the race). Each state representation in the House convenes to decide for whom to cast the single ballot they are given to represent that state. The candidate with a simple majority of all the state ballots (i.e., 26 votes) wins, and at least 34 votes must be cast (the quorum requirement being that 2/3 of the states must be represented).

This procedure was only applied once, in 1824, not due to a tie but to the multiplicity of candidates: Andrew Jackson received 99 votes, short of the 131 then required to be elected, John Quincy Adams 84, William Crawford 41 and Henry Clay 37. Clay being fourth in line was excluded from the Twelfth Amendment procedure, and Adam subsequently won the House vote with 13 out of then 23 state ballots.

Another point worth noting is that electors in the College cast two separate ballots: one for the President, one for the Vice-President. A residency requirement prevents an elector from casting both votes for presidential and vice-presidential candidates both residing in the elector's own state. In practice this residency requirement is not an issue since presidential and vice-presidential candidates on each ticket reside in different states, precisely to avoid this problem. (In the 2000 election, a controversy arose around then vice-presidential candidate Dick Cheney having moved from Texas to Wyoming a few months before the election, so as to avoid a residency conflict with running mate George W. Bush. The case was dismissed by the courts.)

One consequence of having electors cast two separate and in a sense independent votes for President and Vice-President is the need for a separate back-up procedure to elect the Vice-President in case the Electoral College is tied or unable to reach a simple majority for that office. The Twelfth Amendment provides that, similarly to the presidential vote tie-breaking procedure, the Senate then elects the Vice-President, each senator casting one vote. The vice-presidential candidates in the race are limited to those having obtained the two highest numbers of votes (this usually means two candidates, unless there is a tie for the second place). In the 2004 election, this most likely would mean that only John Edwards and Dick Cheney would be in the race. The vice-presidential candidate with a number of votes equal to or greater than the simple majority of all senators (i.e., with 51 votes or more) wins, with the president of the Senate being denied his or her tie-breaking privilege for that election. (A quorum of 2/3 is here again required, i.e. 67 voting senators.)

Finally, the Twentieth Amendment provides for the case in which the procedure laid out by the Twelfth Amendment is itself deadlocked -- which is not unlikely in the House procedure to elect the President, since a number of state representations are currently evenly split Democrats and Republicans. According to the Twentieth Amendment, if a Vice-President has been elected, he or she shall act as President until a President has been elected. If no Vice-President has been elected, Congress may legislate to fill in both offices.

In 2004, it is likely that a tie in the Electoral College would lead to the election of George W. Bush by the House, and it is probable although less certain that the Vice-President elected by the Senate would be Dick Cheney. The possibility of having George W. Bush elected President and John Edward elected Vice-President is very remote but cannot be entirely ruled out given the current composition of the Senate.