Simulating The Electoral College

Adapted from slides by David Meyer
Who will win the presidential election?

- Inputs: Numerous poll results taken in all the states
- Impossible to predict the precise outcome
  - Will Roy Moore be elected senator?
  - Will it rain tomorrow?
- Output: Pr[Clinton wins]
How to win a presidential election

- Need a majority of electoral votes
  - Total = 538 = Representatives (435) + Senators (100) + 3 for DC
  - $538 / 2 = 269$, so to win, need at least 270

- Most states winner-take-all
  - Maine & Nebraska are exceptions, dividing up their electoral votes by congressional districts
Who will win the presidential election?

- Inputs: Numerous poll results taken in all the states
- Impossible to predict the precise outcome
  - Will Roy Moore be elected senator?
  - Will it rain tomorrow?
- Output: $\Pr[\text{Clinton wins}] = \Pr[C \geq 270]$
Simplified story

- Clinton has constant probability $p$ of winning in each state

$$\Pr(C \geq 26) = \sum_{k=26}^{51} \binom{51}{k} p^k (1 - p)^{51-k}$$

- Hard to calculate exactly, but not so hard to simulate!
Simulation

- For each state $s$, draw a random number $n \in [0, 1]$
- If $n < p$, Clinton wins $s$: i.e., set $X_s = 1$
- Clinton wins overall if $\sum_s X_s \geq 26$
- Repeat many (e.g., $10^5$) times

<table>
<thead>
<tr>
<th>$p$</th>
<th>Exact</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$1.982 \times 10^{-13}$</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>$8.129 \times 10^{-7}$</td>
<td>0</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0014</td>
<td>0.0014</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0735</td>
<td>0.0735</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.4970</td>
</tr>
</tbody>
</table>
Less simplified story

- Simplified story
  \[
  \Pr(C \geq 26) = \sum_{k=26}^{51} \binom{51}{k} p^k (1 - p)^{51-k}
  \]

- Closer to the real deal
  \[
  \Pr(C \geq 270) = \sum_{t=270}^{538} \sum_{\sum v_s = t} \prod_{s \in [51]} p_s \prod_{s \notin S} (1 - p_s)
  \]
Simulation

- Let $EV_s$ be the number of electoral votes ascribed to state $s$
- For each state $s$, draw a random number $n \in [0, 1]$
- If $n < p_s$, Clinton wins $s$: i.e., set $X_s = 1$
- Clinton wins overall if $\sum_s (X_s)(EV_s) \geq 270$
- Repeat many (e.g., $10^5$) times
Simulation

- Let $EV_s$ be the number of electoral votes ascribed to state $s$
- For each state $s$, draw a random number $n \in [0, 1]$
- If $n < p_s$, Clinton wins $s$: i.e., set $X_s = 1$
- Clinton wins overall if $\sum_s (X_s)(EV_s) \geq 270$
- Repeat many (e.g., $10^5$) times

Problem with this approach: $p_s \neq 1 - p_s$

What about Johnson, Stein, etc.?
Estimating \( p_s \) from Polling Data

- October 30, 2016 Mitchell Research & Communications poll
  - 953 likely voters in Michigan
  - Clinton: 47%, Trump: 41%, Johnson: 6%, Stein: 2%
  - Margin of error 3.2%
- Assume % Clinton is normal: \( N(\mu_C = 47, \sigma_C = 3.2) \)
- Assume % Trump is normal: \( N(\mu_T = 41, \sigma_T = 3.2) \)
- % Difference is also normal: \( N(\mu_{C-T}, \sqrt{\sigma^2_C + \sigma^2_T}) \)
  - These are the parameters, assuming the percentages are independent.
  - We adjust the standard deviation because they are not:
    - \( N(47 - 41, f \sqrt{3.2^2 + 3.2^2}), f \in \{1, 2, 4\} \)
Estimating $p_s$ from Polling Data, cont’d

- Clinton wins state $s$ if $D_s > 0$

$$p_s = \Pr(\text{Clinton wins state } s) = \frac{1}{\sqrt{2\pi\sigma}} \int_{0}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
66% chance of a Republican victory

HIGHER PROBABILITY
84% chance of a Republican victory
Simulation Results

- Compute $p_s$ as described for each state
  - Treat each of Maine/Nebraska’s districts as its own state for a total of 56 states
- Simulate as described using these more sensible estimates
- $E[C] \approx 299$ and $Pr[C \geq 270] \approx 0.778$
538’s results

- Doesn’t use only the most recent polls. Regresses on past polls to predict future.
  - Also weighs polls by reliability, recency, sample size, etc.
- Pays attention to correlations among states
Extras
An Interesting Aside

- The electoral college has the effect of stretching mid-range probabilities
- Close races become less close as a result