Campaign Ads and Bayesianism

GOV 1347 Lab: Week VI

Matthew E. Dardet

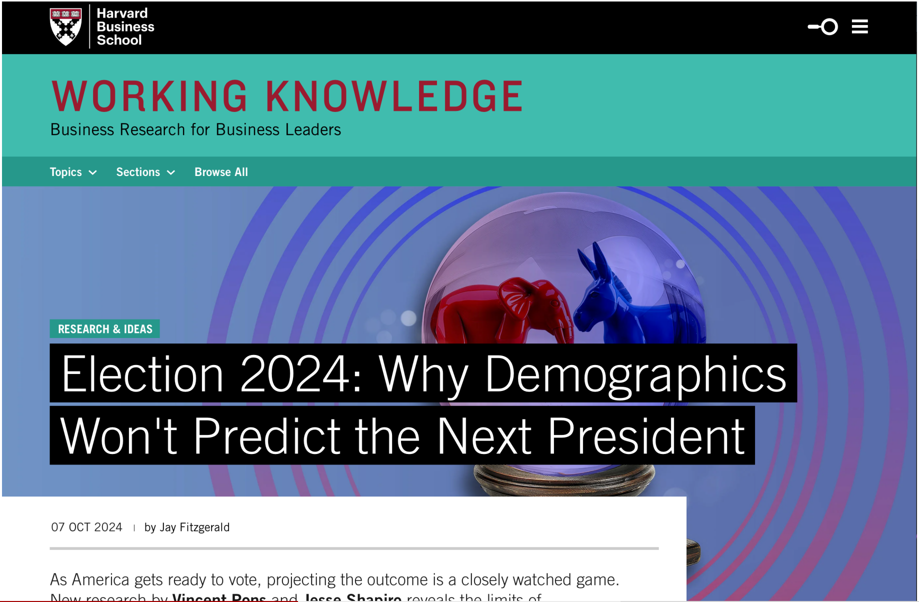
Harvard University

October 9, 2024

Check-In

- Any questions about the course so far?

Hot Off the Press?


Harvard Business School

WORKING KNOWLEDGE

Business Research for Business Leaders

Topics Sections Browse All

RESEARCH & IDEAS

Election 2024: Why Demographics Won't Predict the Next President

07 OCT 2024 | by Jay Fitzgerald

As America gets ready to vote, projecting the outcome is a closely watched game. New research by **Vincent Pons** and **Jesse Shapiro** reveals the limits of

Calvo, Pons, & Shapiro, 2024

Pitfalls of Demographic Forecasts of US Elections
Richard Calvo, Vincent Pons, and Jesse M. Shapiro
NBER Working Paper No. 33016
October 2024
JEL No. C53, D72, J11, P0

ABSTRACT

Many observers have forecast large partisan shifts in the US electorate based on demographic trends. Such forecasts are appealing because demographic trends are often predictable even over long horizons. We backtest demographic forecasts using data on US elections since 1952. We envision a forecaster who fits a model using data from a given election and uses that model, in tandem with a projection of demographic trends, to predict future elections. Even a forecaster with perfect knowledge of future demographic trends would have performed poorly over this period—worse even than one who simply guesses that each election will have a 50-50 partisan split. Enriching the set of demographics available does not change this conclusion. We discuss both mechanical and economic reasons for this finding, and show suggestive evidence that parties adjust their platforms in accordance with changes in the electorate.

Richard Calvo
University of California Berkeley
richard_calvo@berkeley.edu

Jesse M. Shapiro
Department of Economics
Harvard University

This Week

1. **Describing Campaign Advertisements and Spending**
2. **Introduction to Bayesian Thinking & Statistics**
3. **Election Prediction via Bayesian Modelling**

Section 1

Campaign Advertisements and Spending

Campaign Ads and Spending

Visualizing spending using data from:

- Wesleyan Media Project — Ads Spending
  - `ad_campaign_2000-2012.csv`
  - `ad_creative_2000-2012.csv`
  - `ads_2020.csv`
  - `facebook_ads_2020.csv`: Link
  - `facebook_ads_biden_2020.csv`
- Federal Election Commission (FEC) — Campaign Spending
  - `FEC_contributions_by_state_2008_2024.csv`: Link

Section 2

Bayesian Thinking & Statistics

Bayesian Statistics: Motivations

- **Human-Behavioral Foundations:** Probability as numerical expression of subjective belief rather than objective reality.

Bayesian Statistics: Motivations

- **Human-Behavioral Foundations:** Probability as numerical expression of subjective belief rather than objective reality.
- **Updating/Learning:** Can update beliefs by blending prior knowledge with data/evidence to get posterior probabilities.

Bayesian Statistics: Motivations

- **Human-Behavioral Foundations:** Probability as numerical expression of subjective belief rather than objective reality.
- **Updating/Learning:** Can update beliefs by blending prior knowledge with data/evidence to get posterior probabilities.
- Flexible modelling with respect to a number of different probability distributions (e.g., contrast with Normality assumption of OLS).

Bayesian Statistics: Motivations

- **Human-Behavioral Foundations:** Probability as numerical expression of subjective belief rather than objective reality.
- **Updating/Learning:** Can update beliefs by blending prior knowledge with data/evidence to get posterior probabilities.
- Flexible modelling with respect to a number of different probability distributions (e.g., contrast with Normality assumption of OLS).
- Quantify uncertainty explicitly.

Bayesian Statistics: Motivations

- **Human-Behavioral Foundations:** Probability as numerical expression of subjective belief rather than objective reality.
- **Updating/Learning:** Can update beliefs by blending prior knowledge with data/evidence to get posterior probabilities.
- Flexible modelling with respect to a number of different probability distributions (e.g., contrast with Normality assumption of OLS).
- Quantify uncertainty explicitly.
- Do not need p-values or large sample sizes.

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of . . .

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of **Frequentist Statistics or Inference**

- **Parameter interpretation**:

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of **Frequentist Statistics or Inference**

- **Parameter interpretation**: Fixed, unknown quantity.

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of **Frequentist Statistics or Inference**

- **Parameter interpretation**: Fixed, unknown quantity.
- **Approach**: Relies on long-run frequencies of data, using repeated sampling.

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of **Frequentist Statistics or Inference**

- **Parameter interpretation**: Fixed, unknown quantity.
- **Approach**: Relies on long-run frequencies of data, using repeated sampling.
- **Uncertainty (Confidence intervals)**:

Frequentist v. Bayesian Statistics

Throughout this course so far, we have been covering prediction models under the framework of **Frequentist Statistics or Inference**

- **Parameter interpretation**: Fixed, unknown quantity.
- **Approach**: Relies on long-run frequencies of data, using repeated sampling.
- **Uncertainty (Confidence intervals)**: Probability refers to the long-run coverage of the interval (e.g., 95% CI means 95% of intervals trap the true parameter value).

Frequentist v. Bayesian Statistics

Bayesian Statistics

- **Parameter interpretation**:

Frequentist v. Bayesian Statistics

Bayesian Statistics

- **Parameter interpretation**: Random variable with a probability distribution.

Frequentist v. Bayesian Statistics

Bayesian Statistics

- **Parameter interpretation**: Random variable with a probability distribution.
- **Approach**: Updates prior beliefs with data evidence using Bayes' Theorem.

Frequentist v. Bayesian Statistics

Bayesian Statistics

- **Parameter interpretation**: Random variable with a probability distribution.
- **Approach**: Updates prior beliefs with data evidence using Bayes' Theorem.
- **Uncertainty (Credible intervals)**: The interval represents the probability that the parameter lies within it (e.g., 95% credible interval means a 95% chance the parameter is in this interval).

Bayes' Theorem

Bayes' Theorem

- Two events: A and B
- From the conditional probability formula, we have:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Bayes' Theorem

- Two events: A and B
- From the conditional probability formula, we have:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Theorem

- Two events: A and B
- From the conditional probability formula, we have:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

$$P(B) = P(B|A)P(A) + P(B|\text{not } A)P(\text{not } A)$$

Bayesian Updating

- Bayesian updating:

$$P(A|B) \propto P(B|A) \times P(A)$$

Bayesian Updating

- Bayesian updating:

$$P(A|B) \propto P(B|A) \times P(A)$$

- **Posterior belief**: $P(A|B)$

Bayesian Updating

- Bayesian updating:

$$P(A|B) \propto P(B|A) \times P(A)$$

- **Posterior belief**: $P(A|B)$
- **Likelihood (data)**: $P(B|A)$

Bayesian Updating

- Bayesian updating:

$$P(A|B) \propto P(B|A) \times P(A)$$

- **Posterior belief**: $P(A|B)$
- **Likelihood (data)**: $P(B|A)$
- **Prior belief**: $P(A)$

Bayes' Rule

Another formulation:

Bayes' Rule

Another formulation: Bayes' Rule helps us update the probability of a hypothesis given new evidence.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Bayes' Rule

Another formulation: Bayes' Rule helps us update the probability of a hypothesis given new evidence.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- **P(H | E)**: Posterior probability, the probability of the hypothesis H given the evidence E.

Bayes' Rule

Another formulation: Bayes' Rule helps us update the probability of a hypothesis given new evidence.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- **P(H | E)**: Posterior probability, the probability of the hypothesis H given the evidence E.
- **P(E | H)**: Likelihood, the probability of the evidence E assuming the hypothesis H is true.

Bayes' Rule

Another formulation: Bayes' Rule helps us update the probability of a hypothesis given new evidence.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- **P(H | E)**: Posterior probability, the probability of the hypothesis H given the evidence E.
- **P(E | H)**: Likelihood, the probability of the evidence E assuming the hypothesis H is true.
- **P(H)**: Prior probability, the initial probability of the hypothesis before seeing the evidence.

Bayes' Rule

Another formulation: Bayes' Rule helps us update the probability of a hypothesis given new evidence.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- **P(H | E)**: Posterior probability, the probability of the hypothesis H given the evidence E.
- **P(E | H)**: Likelihood, the probability of the evidence E assuming the hypothesis H is true.
- **P(H)**: Prior probability, the initial probability of the hypothesis before seeing the evidence.
- **P(E)**: Marginal likelihood, the total probability of the evidence under all possible hypotheses.

Bayes' Rule Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Bayes' Rule Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- Imagine you want to know the probability someone has a disease (H) given a positive test result (E).

Bayes' Rule Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- Imagine you want to know the probability someone has a disease (H) given a positive test result (E).
- **P(H)**: Prior, probability someone has the disease based on population data.

Bayes' Rule Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- Imagine you want to know the probability someone has a disease (H) given a positive test result (E).
- **P(H)**: Prior, probability someone has the disease based on population data.
- **P(E | H)**: Likelihood, probability of a positive test if they have the disease.

Bayes' Rule Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- Imagine you want to know the probability someone has a disease (H) given a positive test result (E).
- **P(H)**: Prior, probability someone has the disease based on population data.
- **P(E | H)**: Likelihood, probability of a positive test if they have the disease.
- **P(E)**: Marginal likelihood, the probability of getting a positive test overall.

Bayes' Rule Example with Numbers

Suppose you are testing for a rare disease:

- The prior probability $P(H)$, or the probability that a randomly selected person has the disease, is 1% (0.01).
- The test has a sensitivity $P(E|H)$ of 90% (0.90), meaning 90% of those with the disease test positive.
- The test has a false positive rate, or $P(E|H^c)$, of 5% (0.05).

Bayes' Rule Example with Numbers

Suppose you are testing for a rare disease:

- The prior probability $P(H)$, or the probability that a randomly selected person has the disease, is 1% (0.01).
- The test has a sensitivity $P(E|H)$ of 90% (0.90), meaning 90% of those with the disease test positive.
- The test has a false positive rate, or $P(E|H^c)$, of 5% (0.05).

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Bayes' Rule Example with Numbers

Step 1: Compute the marginal likelihood $P(E)$

$$P(E) = P(E|H)P(H) + P(E|H^c)P(H^c)$$

Bayes' Rule Example with Numbers

Step 1: Compute the marginal likelihood $P(E)$

$$P(E) = P(E|H)P(H) + P(E|H^c)P(H^c)$$

- $P(E|H) = 0.90$ (True Positive Rate)
- $P(H) = 0.01$ (Prior Probability)
- $P(E|H^c) = 0.05$ (False Positive Rate)
- $P(H^c) = 1 - P(H) = 0.99$ (Probability of not having the disease)

Bayes' Rule Example with Numbers

Step 1: Compute the marginal likelihood $P(E)$

$$P(E) = P(E|H)P(H) + P(E|H^c)P(H^c)$$

- $P(E|H) = 0.90$ (True Positive Rate)
- $P(H) = 0.01$ (Prior Probability)
- $P(E|H^c) = 0.05$ (False Positive Rate)
- $P(H^c) = 1 - P(H) = 0.99$ (Probability of not having the disease)

$$P(E) = (0.90)(0.01) + (0.05)(0.99) = 0.009 + 0.0495 = 0.0585$$

Bayes' Rule Example with Numbers

Step 2: Update the posterior probability $P(H|E)$

$$P(H|E) = \frac{(0.90)(0.01)}{0.0585} = \frac{0.009}{0.0585} \approx 0.154$$

Bayes' Rule Example with Numbers

Step 2: Update the posterior probability $P(H|E)$

$$P(H|E) = \frac{(0.90)(0.01)}{0.0585} = \frac{0.009}{0.0585} \approx 0.154$$

- The posterior probability that a person has the disease, given a positive test result, is approximately **15.4%**.

Bayesian Inference

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

Bayesian Inference

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

- **Posterior**: $p(\theta|Y)$

Bayesian Inference

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

- **Posterior**: $p(\theta|Y)$
- **Likelihood**: $p(Y|\theta)$

Bayesian Inference

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

- **Posterior**: $p(\theta|Y)$
- **Likelihood**: $p(Y|\theta)$
- **Prior**: $p(\theta)$

Bayesian Inference

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

- **Posterior**: $p(\theta|Y)$
- **Likelihood**: $p(Y|\theta)$
- **Prior**: $p(\theta)$
- **Marginal likelihood**: $\int p(Y|\theta)p(\theta)d\theta$

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters
2. Prior: one's belief about θ before seeing the data

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters
2. Prior: one's belief about θ before seeing the data

- **Posterior**: one's belief about θ after seeing the data

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters
2. Prior: one's belief about θ before seeing the data

- **Posterior**: one's belief about θ after seeing the data

Bayesian prediction:

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters
2. Prior: one's belief about θ before seeing the data

- **Posterior**: one's belief about θ after seeing the data

Bayesian prediction:

- **Prior predictive distribution**:

$$p(Y) = \int p(Y|\theta)p(\theta)d\theta$$

Bayesian Model

Bayesian model:

1. Likelihood: data generating process given parameters
2. Prior: one's belief about θ before seeing the data

- **Posterior**: one's belief about θ after seeing the data

Bayesian prediction:

- **Prior predictive distribution**:

$$p(Y) = \int p(Y|\theta)p(\theta)d\theta$$

- **Posterior predictive distribution**:

$$p(Y^{\text{new}}|Y) = \int p(Y^{\text{new}}|\theta)p(\theta|Y)d\theta$$

Many Frequentist Concepts Have Bayesian Analogues

Example: Linear Regression

Frequentist: $Y = X\beta + \epsilon$, where

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta)^2$$

Bayesian Linear Regression

- In Bayesian linear regression, we treat the model parameters as random variables with probability distributions.

Bayesian Linear Regression

- In Bayesian linear regression, we treat the model parameters as random variables with probability distributions.

Model:

$$y = X\beta + \epsilon$$

Bayesian Linear Regression

Bayesian Approach:

- **Prior Distribution**: Assigns a prior belief to the parameters, e.g., $\beta \sim N(0, \sigma^2 I)$.

$$\beta \sim N(0, \sigma^2 I)$$

- **Likelihood**: The probability of the observed data given the parameters.

$$y|X, \beta \sim N(X\beta, \sigma^2)$$

Bayesian Linear Regression

Bayesian Approach:

- **Prior Distribution**: Assigns a prior belief to the parameters, e.g., $\beta \sim N(0, \sigma^2 I)$.

$$\beta \sim N(0, \sigma^2 I)$$

- **Likelihood**: The probability of the observed data given the parameters.

$$y|X, \beta \sim N(X\beta, \sigma^2)$$

- **Posterior Distribution**: Combines the prior and likelihood using Bayes' theorem to update beliefs about β given the data.

$$p(\beta|y, X) \propto p(y|X, \beta)p(\beta)$$

Bayesian Linear Regression

Let's compare frequentist and Bayesian linear regression in `R`...

Using RStan.

Many Frequentist Concepts Have Bayesian Analogues

Example: Regularization

Hans (2011) overviews how to implement Bayesian elastic-net.

Many Frequentist Concepts Have Bayesian Analogues

Example: Regularization

Hans (2011) overviews how to implement Bayesian elastic-net.

- Regression parameter estimates based on regularization procedures can be interpeted as mode of posterior distribution when likelihood component is normal regression model.

Many Frequentist Concepts Have Bayesian Analogues

Example: Regularization

Hans (2011) overviews how to implement Bayesian elastic-net.

- Regression parameter estimates based on regularization procedures can be interpeted as mode of posterior distribution when likelihood component is normal regression model.
- Elastic-net penalty can be expressed as unnormalized prior density:

$$p(\beta|\lambda, \alpha) \propto \exp\left(-\lambda\left(\alpha|\beta|^2 + (1-\alpha)|\beta|_1\right)\right)$$

Many Frequentist Concepts Have Bayesian Analogues

Example: Regularization

Hans (2011) overviews how to implement Bayesian elastic-net.

- Regression parameter estimates based on regularization procedures can be interpeted as mode of posterior distribution when likelihood component is normal regression model.
- Elastic-net penalty can be expressed as unnormalized prior density:

$$p(\beta|\lambda, \alpha) \propto \exp\left(-\lambda\left(\alpha|\beta|^2 + (1-\alpha)|\beta|_1\right)\right)$$

- Referred to as Bayesian shrinkage.

Section 3

Bayesian Election Prediction

The Economists's Model

Putting the pieces together

After making all of these adjustments to polls' reported results, we are ready to use them to update our prior. Our method is an expansion of a technique first published by Drew Linzer, a political scientist, in 2013. It uses a statistical technique called Markov Chain Monte Carlo (MCMC), which explores thousands of different values for each parameter in our model, and evaluates both how well they explain the patterns in the data and how plausible they are given the expectations from our prior. For example, what would the election look like if all online pollsters over-estimated the Republicans' vote share by five percentage points? How about if all national polls over-estimated Democrats by two? If state polls of Michigan are oscillating by ten percentage points at a time, the model will incorporate more uncertainty in its prediction of the vote there—and in its predictions of the vote in similar states, such as Ohio.

FiveThirtyEight's Model

More recently, political scientist/survey statistician/pollster Drew Linzer combined aspects of these and other approaches in a Bayesian dynamic linear model of state and national polls of the 2012 general election. In 2016, Natalie Jackson at The Huffington Post also worked on a statistical forecasting model that ran on the Markov chain. And, as linked above, I worked with Gelman and statistician Merlin Heidemanns to add additional poll-level adjustments as well as other factors to model the 2020 general election in this way.

Linzer's Dynamic Baysian Forecasting Model

- Linzer's 2013 model combines historical forecasting with real-time state-level polling data.

Linzer's Dynamic Baysian Forecasting Model

- Linzer's 2013 model combines historical forecasting with real-time state-level polling data.
- "Sequence of state-level preelection polls can be used to estimate both current voter preferences and forecasts of election outcome for every state and on every day of the campaign, regardless of whether a survey was conducted on that day."

Linzer's Dynamic Baysian Forecasting Model

- Linzer's 2013 model combines historical forecasting with real-time state-level polling data.
- "Sequence of state-level preelection polls can be used to estimate both current voter preferences and forecasts of election outcome for every state and on every day of the campaign, regardless of whether a survey was conducted on that day."

Key Features of Linzer's Model

- **Hierarchical structure:** "borrows strength" across states and time.

Key Features of Linzer's Model

- **Hierarchical structure:** "borrows strength" across states and time.
- **Random-walk priors:** Smooth estimates over time and filter out day-to-day noise in the polls.

Key Features of Linzer's Model

- **Hierarchical structure:** "borrows strength" across states and time.
- **Random-walk priors:** Smooth estimates over time and filter out day-to-day noise in the polls.
- **Daily tracking:** Provides continuous updates of voter preferences at both the state and national levels.

Key Features of Linzer's Model

- **Hierarchical structure:** "borrows strength" across states and time.
- **Random-walk priors:** Smooth estimates over time and filter out day-to-day noise in the polls.
- **Daily tracking:** Provides continuous updates of voter preferences at both the state and national levels.
- **Electoral College Simulation:** Estimates the candidates' probability of winning a majority of votes.

The Dynamic Bayesian Approach

- Bayesian framework combines **prior information** with new data (polls) using:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

- Initial state-level vote shares are predicted from historical data.

The Dynamic Bayesian Approach

- Bayesian framework combines **prior information** with new data (polls) using:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

- Initial state-level vote shares are predicted from historical data.
- As more polls are released, the posterior distribution is updated to refine state-level forecasts.

The Dynamic Bayesian Approach

- Bayesian framework combines **prior information** with new data (polls) using:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

- Initial state-level vote shares are predicted from historical data.
- As more polls are released, the posterior distribution is updated to refine state-level forecasts.
- Polling gaps are filled using hierarchical modeling, allowing the model to "borrow strength" across states.

The Dynamic Bayesian Approach

- Want to model number of respondents who support Democratic candidate y_k:

$$y_k \sim \text{Binomial}(\pi_{i[k]j[k]}, n_k)$$

The Dynamic Bayesian Approach

- Want to model number of respondents who support Democratic candidate y_k:

$$y_k \sim \text{Binomial}(\pi_{i[k]j[k]}, n_k)$$

- π_{ij}, the proportion of voters in state i on day j who would tell pollsters they intend to vote for Democrat. k indexes individual polls.

The Dynamic Bayesian Approach

- Want to model number of respondents who support Democratic candidate y_k:

$$y_k \sim \text{Binomial}(\pi_{i[k]j[k]}, n_k)$$

- π_{ij}, the proportion of voters in state i on day j who would tell pollsters they intend to vote for Democrat. k indexes individual polls.

$$\pi_{ij} = \text{logit}^{-1}(\beta_{ij} + \delta_j)$$

  - β_{ij}: State-level effect that captures long-term dynamics of voter preferences in state i
  - δ_j: National-level effect that detects systematic departures from β_{ij} on day j due to short-term campaign factors that act as "uniform swing" across all states.

Random Walks and Markov Chains

- Random walks are a form of stochastic process that consists of “random steps” on some space (e.g., stock prices).
  - Stochastic processes generally describe random variables in probability space indexed by time.

Random Walks and Markov Chains

- Random walks are a form of stochastic process that consists of "random steps" on some space (e.g., stock prices).
  - Stochastic processes generally describe random variables in probability space indexed by time.
- Each step depends on previous step but with randomness (stochasticity).

Random Walks and Markov Chains

- Random walks are a form of stochastic process that consists of "random steps" on some space (e.g., stock prices).
  - Stochastic processes generally describe random variables in probability space indexed by time.
- Each step depends on previous step but with randomness (stochasticity).
- Markov chains are stochastic processes where the next step only depends on the current state. A simple (1D) random walk is an example of Markov chain.

Random Walks and Markov Chains

- Random walks are a form of stochastic process that consists of "random steps" on some space (e.g., stock prices).
  - Stochastic processes generally describe random variables in probability space indexed by time.
- Each step depends on previous step but with randomness (stochasticity).
- Markov chains are stochastic processes where the next step only depends on the current state. A simple (1D) random walk is an example of Markov chain.

Handling Poll Variability with Random Walk Priors

- Polls fluctuate due to sampling error and campaign events.

Handling Poll Variability with Random Walk Priors

- Polls fluctuate due to sampling error and campaign events.
- Linzer's model smooths these variations using random-walk priors for the state and national-level effects.

Handling Poll Variability with Random Walk Priors

- Polls fluctuate due to sampling error and campaign events.
- Linzer's model smooths these variations using random-walk priors for the state and national-level effects.
- Start with informative Normal prior distribution based on state-level historical forecasts h_i:

$$\beta_{iJ} \sim N(\text{logit}(h_i), s_i^2)$$

Handling Poll Variability with Random Walk Priors

The dynamic nature of β_{ij} and δ_j is modeled as:

$$\beta_{ij}|\beta_{i,j+1} \sim N(\beta_{i,j+1}, \sigma_\beta^2)$$

Where the variance σ_β^2 captures the rate of daily change in β_{ij}.

Handling Poll Variability with Random Walk Priors

The dynamic nature of β_{ij} and δ_j is modeled as:

$$\beta_{ij}|\beta_{i,j+1} \sim N(\beta_{i,j+1}, \sigma_\beta^2)$$

Where the variance σ_β^2 captures the rate of daily change in β_{ij}.

Likewise, for δ_j:

$$\delta_j|\delta_{j+1} \sim N(\delta_{j+1}, \sigma_\delta^2)$$

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.
- **Markov Chain Monte Carlo (MCMC)** represents a general class of algorithms that construct a Markov chain whose stationary distribution is the target distribution.

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.
- **Markov Chain Monte Carlo (MCMC)** represents a general class of algorithms that construct a Markov chain whose stationary distribution is the target distribution.
  - Markov Chain $\rightarrow$ distribution of sampled draws depends on last value drawn.

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.
- **Markov Chain Monte Carlo (MCMC)** represents a general class of algorithms that construct a Markov chain whose stationary distribution is the target distribution.
  - Markov Chain $\rightarrow$ distribution of sampled draws depends on last value drawn.
  - Monte Carlo $\rightarrow$ simulation/sampling.

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.
- **Markov Chain Monte Carlo (MCMC)** represents a general class of algorithms that construct a Markov chain whose stationary distribution is the target distribution.
  - Markov Chain $\rightarrow$ distribution of sampled draws depends on last value drawn.
  - Monte Carlo $\rightarrow$ simulation/sampling.
- Approximate distributions are improved at each step in simulation $\rightarrow$ converge to target distribution.
- MCMC is widely used in Bayesian inference to approximate posterior distributions when analytical solutions are difficult or impossible.

Markov Chain Monte Carlo (MCMC)

- Posterior distributions can be very complex and make sampling difficult.
- **Markov Chain Monte Carlo (MCMC)** represents a general class of algorithms that construct a Markov chain whose stationary distribution is the target distribution.
  - Markov Chain $\rightarrow$ distribution of sampled draws depends on last value drawn.
  - Monte Carlo $\rightarrow$ simulation/sampling.
- Approximate distributions are improved at each step in simulation $\rightarrow$ converge to target distribution.
- MCMC is widely used in Bayesian inference to approximate posterior distributions when analytical solutions are difficult or impossible.
- MCMC allows us to estimate expected values, posterior distributions, and other statistical quantities through iterative sampling.

Gibbs Sampler

- **Gibbs Sampler** is a specific MCMC algorithm.
- Samples from the conditional distribution of each variable, one at a time, while keeping all other variables fixed.

Gibbs Sampler

- **Gibbs Sampler** is a specific MCMC algorithm.
- Samples from the conditional distribution of each variable, one at a time, while keeping all other variables fixed.
- For parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$, sample each component from its full conditional distribution:

$$\theta_1 \sim p(\theta_1 | \theta_2, \theta_3, ..., \theta_k)$$

$$\theta_2 \sim p(\theta_2 | \theta_1, \theta_3, ..., \theta_k)$$

$$\vdots$$

Gibbs Sampler

- **Gibbs Sampler** is a specific MCMC algorithm.
- Samples from the conditional distribution of each variable, one at a time, while keeping all other variables fixed.
- For parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$, sample each component from its full conditional distribution:

$$\theta_1 \sim p(\theta_1|\theta_2, \theta_3, ..., \theta_k)$$

$$\theta_2 \sim p(\theta_2|\theta_1, \theta_3, ..., \theta_k)$$

$$\vdots$$

- Gibbs sampler is useful when these full conditionals are easy to sample from.

Putting It All Together

- Basically, we model each state's Democratic vote share as iterative functions of state-level and national-level trends, with the latter shrinking in importance as we get more polls closer to election date.

Putting It All Together

- Basically, we model each state's Democratic vote share as iterative functions of state-level and national-level trends, with the latter shrinking in importance as we get more polls closer to election date.
- Bayesian elements incorporate differing levels of uncertainty in the model, allowing us to make probabilistic predictions about election outcomes at any date in the election cycle using polls and historical forecasts.

Putting It All Together

- Basically, we model each state's Democratic vote share as iterative functions of state-level and national-level trends, with the latter shrinking in importance as we get more polls closer to election date.
- Bayesian elements incorporate differing levels of uncertainty in the model, allowing us to make probabilistic predictions about election outcomes at any date in the election cycle using polls and historical forecasts.

Linzer (2013) uses this Bayesian model and estimates it using MCMC in a program called WinBUGS and `R`. Now we generally use `STAN` and `R`.

Blog Extensions (Optional)

1. **Campaign Spending.** How much have campaigns spent in 2024? How can campaign spending be used to predict the election outcomes?
2. **Social Media.** How much do campaigns spend on social media ads? Does social media influence election outcomes? How will this spending influence the 2024 elections? Many rich descriptive opportunities for this week!
3. **Bayesian/Probabilistic Election Prediction.** Implement a Bayesian model for your 2024 election prediction using the Linzer (2013) model or another probabilistic model such as the Binomial logistic model and explanations of *The Economist* or *FiveThirtyEight*'s Bayesian election models as guides.
   - Kremp's Implementation: https://www.slate.com/features/pkremp_forecast/report.html
   - Morris' Implementation: https://github.com/TheEconomist/us-potus-model/tree/master