

Um Método Bayesiano para agregar pesquisas eleitorais no Brasil

Guilherme Jardim Duarte

Jota

guilherme.duarte@jota.info

August 9, 2018

Motivação

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

- Pesquisas eleitorais indicam resultados distintos: diferenças entre metodologia, erro amostral
- Em qual confiar?
 - Verificar acurácia dos institutos
 - Agregar pesquisas de institutos diferentes
- Agregar pesquisas pode melhorar capacidade preditiva:

$$MSE = Bias^2 + Var$$

- Inferência Bayesiana é uma maneira straightforward de agregar

Pesquisas eleitorais - como funcionam?

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Objetivo: fazer inferência em relação ao apoio em relação aos candidatos ou pré-candidatos. Ou ao resultado da eleição.
Amostragem: representar posição da população com base em uma quantidade limitada de indivíduos.



Emir Sader
@emirsader

Seguir

Como uma mostra de 2884 pessoas pode espelhar um universo de 140 milhões de pessoas, como faz o Datafalha?

14:24 - 29 de ago de 2014

Métodos realizados no Brasil:

- Amostragem aleatória simples: não realizado
- Amostragem por cotas (Domicílios x Pontos de fluxo)
- Amostragem aleatória de números de telefone

Problemas com as pesquisas

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Margem de erro: como calcular?

Erro amostral x outros tipos de erros: *total survey error*

- *frame bias*
- não-resposta
- erro de medida
- erro de especificação
- *shifting attitudes*

Melhorando previsão eleitoral - *poll aggregation*

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Combinar informações provenientes de diversas pesquisas eleitorais geralmente é uma boa ideia

- Resultados mais acurados
- Melhora erro padrão
- Facilita comparação entre mais de um instituto

Todavia:

- Institutos não são diretamente comparáveis. *House effects*
- Qual modelo estatístico consegue realizar essa comparação?

Poll aggregation nos Estados Unidos

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

É bastante comum e ocorre há muito tempo.

WANG, Samuel. *Origins of Presidential poll aggregation: A perspective from 2004 to 2012*. International Journal of Forecasting. 2015.

Alguns pioneiros:

- RealClearPolitics (2002)
- Andrew Tanenbaum's Electoralvote.com (2004)

Modelo do Nate Silver

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Nate Silver se tornou famoso ao prever com sucesso o resultado das eleições presidenciais dos EUA nos 50 estados e DC em 2012.

Método (<https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeight-2016-general-election-forecast/>):

- *polls-only, polls-plus, now-cast*
- Coletar dados. Exceção: *ban list*
- Por estado, média ponderada conforme ranking dos Institutos. Considerado o tamanho da amostra e se a pesquisa é recente.
- Ajustes: *house effect, loess regression*, entre outros.
- Ajuste para calcular probabilidades para o candidato (Simulação).

Probabilidade nas eleições americanas

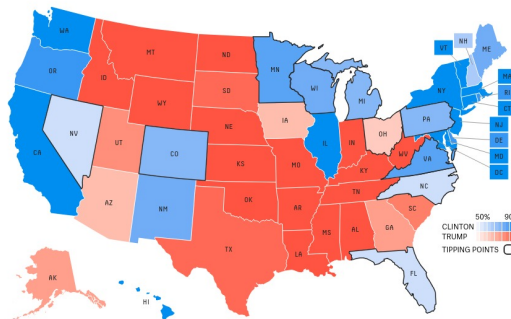
Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Who will win the presidency?



Chance of winning



Probabilidade nas eleições americanas

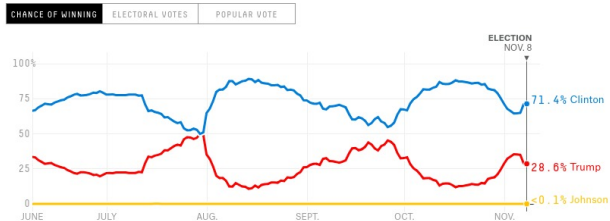
Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Problema de se reportar probabilidade em vez da porcentagem do candidato.

How the forecast has changed

We'll be updating our forecasts every time new data is available, every day through Nov. 8.



Como agregar pesquisas no Brasil?

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Poucas Pesquisas. Inferência Bayesiana. Inspiração: Drew Linzer, Simon Jackman, Jeff Lewis

Inferências Bayesiana -> mais adequada

Modelo bem definido

Custos: MCMC é uma rotina pesada

O que é inferência bayesiana? Verossimilhança

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Máxima verossimilhança: Definimos uma função de verossimilhança e a maximizamos.

Função de verossimilhança pode ser entendida como a probabilidade dos dados dado um parâmetro.

$$F(\theta) = P(X|\theta)$$

Qual é o θ que descreve uma distribuição binomial para um moeda? Nos lançamentos, saíram 12 caras e 16 coroas

O que é inferência bayesiana? Priors e Posteriors

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Podemos obter a probabilidade de cada parâmetro, $P(\theta|X)$, em vez de $P(X|\theta)$.

$$P(\theta|X) = \frac{P(\theta|X) * P(X)}{P(\theta)}$$

A mesma expressão:

$$\textit{posterior} = \textit{verossimilhanca} * \textit{prior}$$

Prior representa a crença em cima dos parâmetros.

Vantagens e desvantagens

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

- Vantagens

- Bayesianismo e subjetivismo guardam coerência lógica
- Facilidade para descrever modelos (Ex: pontos ideais, IRT, eleições, multinível)

- Desvantagens

- Custo computacional. Métodos: Metrópolis, MCMC, HMC

Método para agregação

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Dados: 2º turno eleições presidenciais no Brasil

Modelo:

- Supomos uma variável latente de intenção de voto, calculada para cada candidato por vez.
- Essa variável muda no tempo. Em $t = 0$, o prior é uma uniforme entre 0 e 1. Em $t \geq 1$, é uma normal com $N(\mu_{t-1}, 0.0025)$
- Cada t representa um dia, com $max(t)$ indicando o dia das eleições.
- Verossimilhança é dada pelos resultados das pesquisas, com variância proporcional ao tamanho da amostra do instituto j

Exercício 1

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

O que ocorreria se apenas simulássemos a intenção de voto sem verossimilhança?

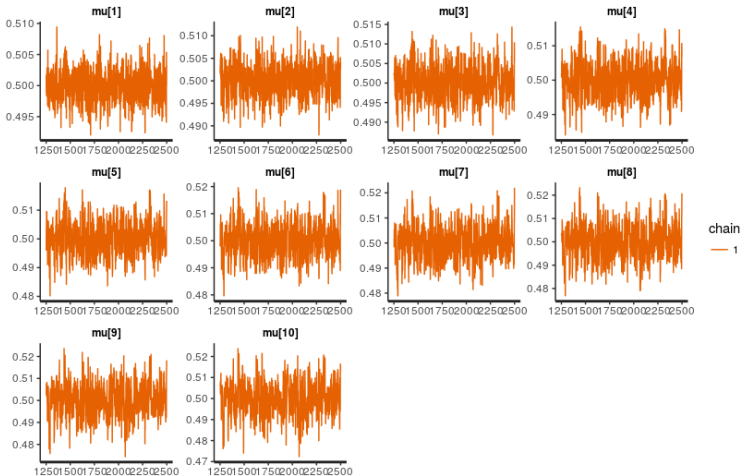
Basta usar um modelo apenas definir os priors.

Aqui definiremos, em $t = 0$, uma prior para $\mu_{t=0}$ como uma $N(0.5, sd = 0.0025)$. As priors para $\mu_{t \geq 1}$ serão $N(\mu_{t-1}, sd = 0.0025)$

Exercício 1 - Convergência dos Parâmetros

Agregação
Bayesiana de
pesquisas no
Brasil

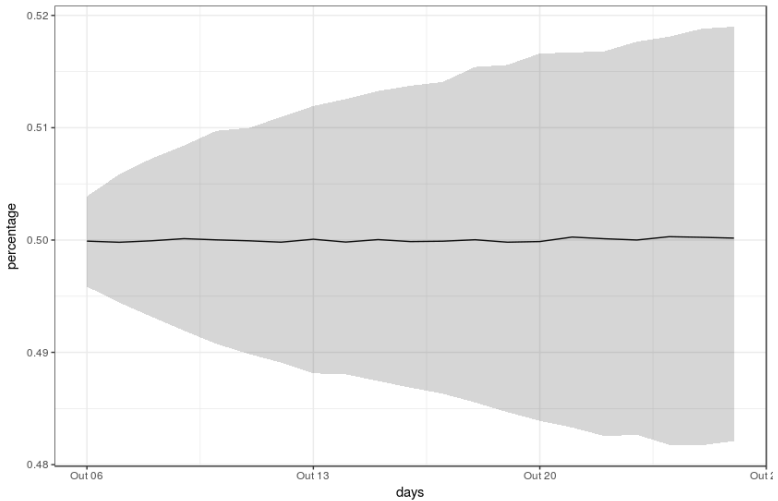
Guilherme
Jardim Duarte



Exercício 1 - Plot

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte



Exercício 2

Agregação
Bayesiana de
pesquisas no
Brasil

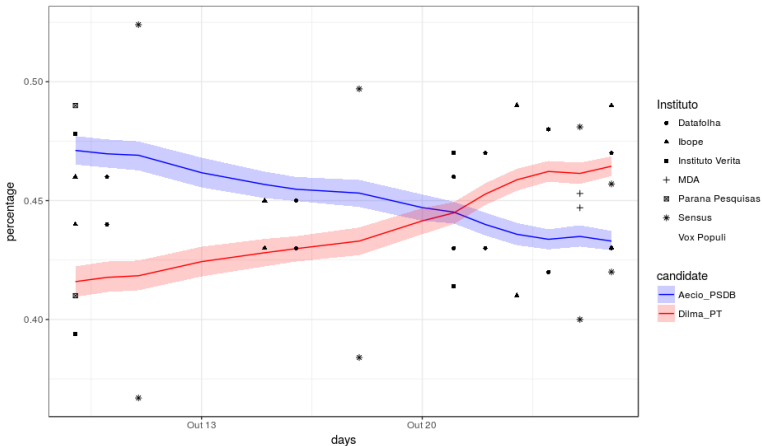
Guilherme
Jardim Duarte

- Inserir verossimilhança no modelo
- Vamos assumir que:
 - 1 Todo instituto tem peso igual
 - 2 Não existem *house effects*

Exercício 2 - Plot

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte



Exercício 3

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

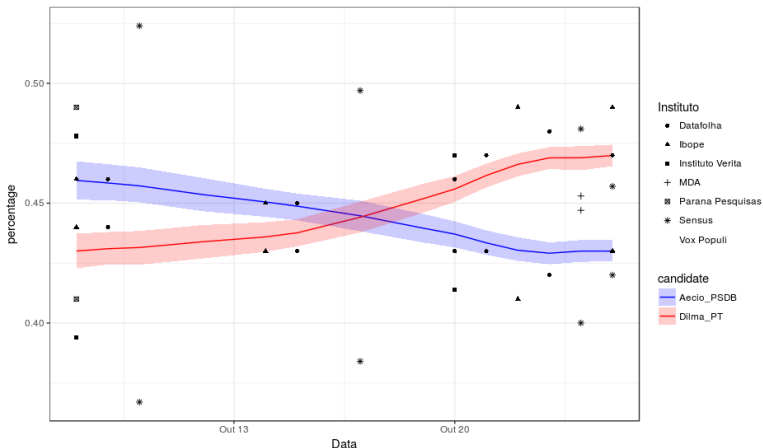
- Vamos dar pesos diferentes para cada instituto.
- faremos isso multiplicando a variancia por um fator.
- No caso presente, daremos maior peso para o Datafolha e para o Ibope ($1 \cdot \text{var}$)
- Para os outros institutos, ($2 \cdot \text{var}$)
- Este ponto é controverso. Nate Silver atribuiu peso a institutos. Outros criticam.
- Não temos uma maneira correta de atribuir pesos no Brasil.

Exercício 3 - Plot

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

Resultado se aproxima das porcentagens obtidas por Dilma (48%) e Aécio (45%).



Exercício 4

Agregação
Bayesiana de
pesquisas no
Brasil

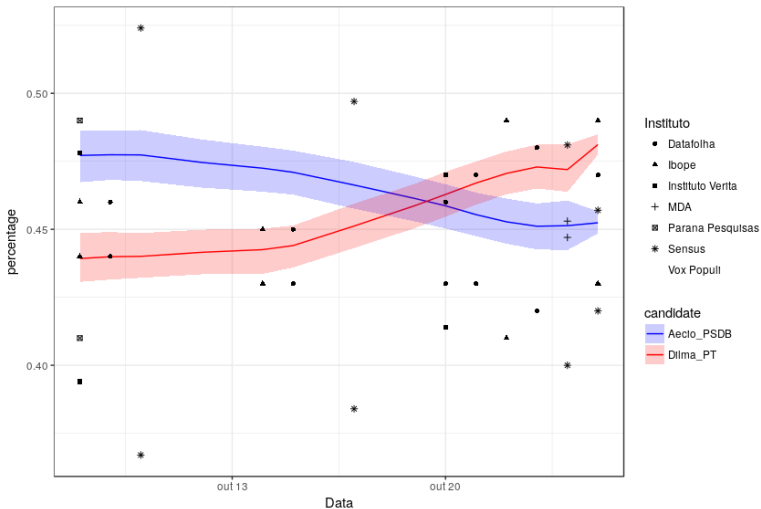
Guilherme
Jardim Duarte

- Vamos estudar *house effects* para cada instituto
- *House effects* significam a tendência de um instituto em favorecer um candidato por conta de sua metodologia
- Para fazer isso, basta inserir um novo parâmetro γ para cada instituto i
- A função de verossimilhança entrará com uma Normal com $N(\mu_t + \gamma_i)$
- Aqui precisamos fazer um dos três tipos de pressuposições suficientes para a estimação:
 - 1 Basear os house effects no resultado final das eleições;
 - 2 Assumir que o efeito de um ou mais institutos é zero;
 - 3 Assumir que a soma de todos os efeitos é zero;
- Escolheremos a primeira.

Exercício 4 - Plot

Agregação
Bayesiana de
pesquisas no
Brasil

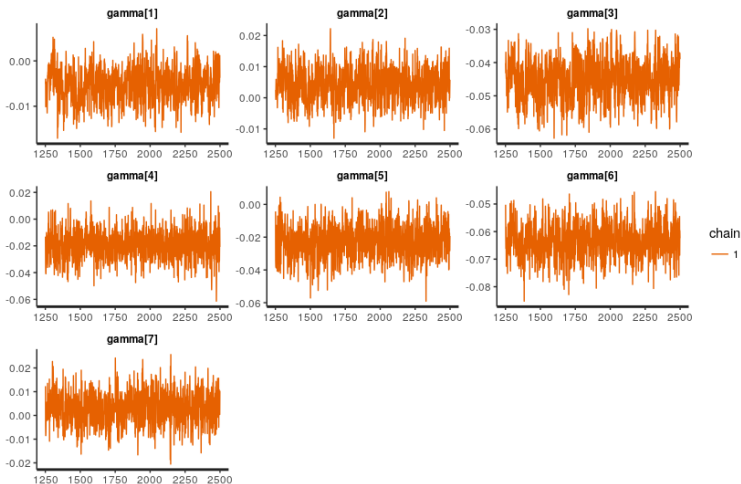
Guilherme
Jardim Duarte



Exercício 4 - Convergência

Agregação
Bayesiana de
pesquisas no
Brasil

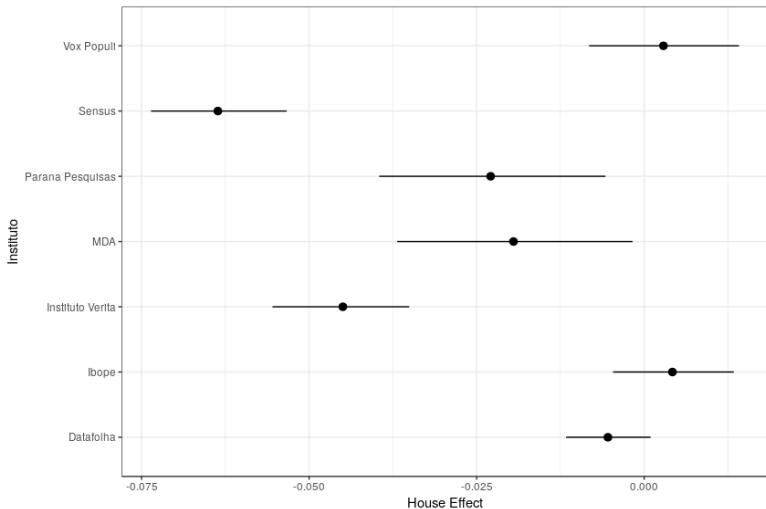
Guilherme
Jardim Duarte



Exercício 4 - House Effects para a candidata Dilma(PT)

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte



Conclusões

Agregação
Bayesiana de
pesquisas no
Brasil

Guilherme
Jardim Duarte

- Inferência Bayesiana é uma ótima forma de agregar pesquisas porque permite que consigamos inserir elementos de uma maneira direta e expressiva.
- Elementos para se acrescentar, por exemplo:
 - 1 Cálculo de probabilidade para cada candidato
 - 2 Inserir erros mais plausíveis. V. Shirani-Mehr, Goel, Rotschild, Gelman, 2018.
 - 3 Calcular house effects com outros pressupostos

Obrigado!

twitter: @guilhermejd1 email: guilherme.duarte@jota.info