Creating structured and flexible models: some open problems

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

26 July 2010

Themes

- *Weakly informative priors* let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data
- *Interaction models* to better learn from the data

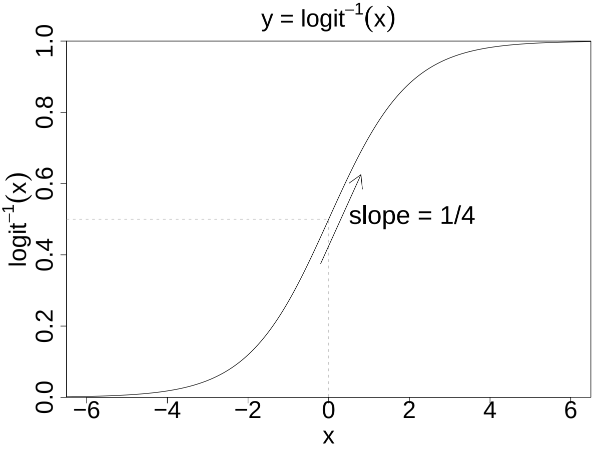
Themes

- *Weakly informative priors* let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data
- *Interaction models* to better learn from the data

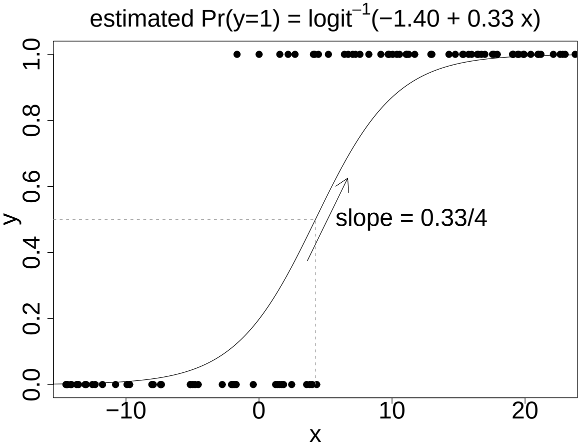
Themes

- *Weakly informative priors* let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data
- *Interaction models* to better learn from the data

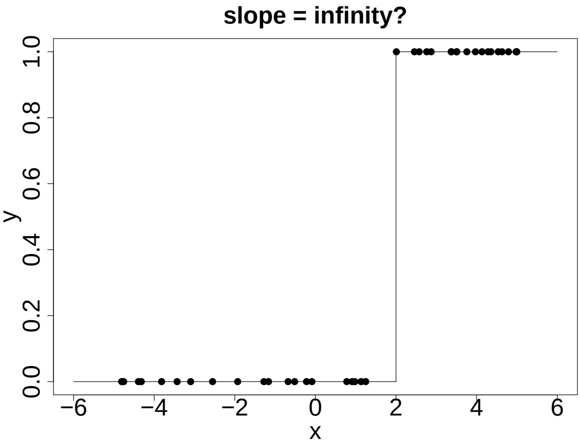
Logistic regression



A clean example



The problem of separation



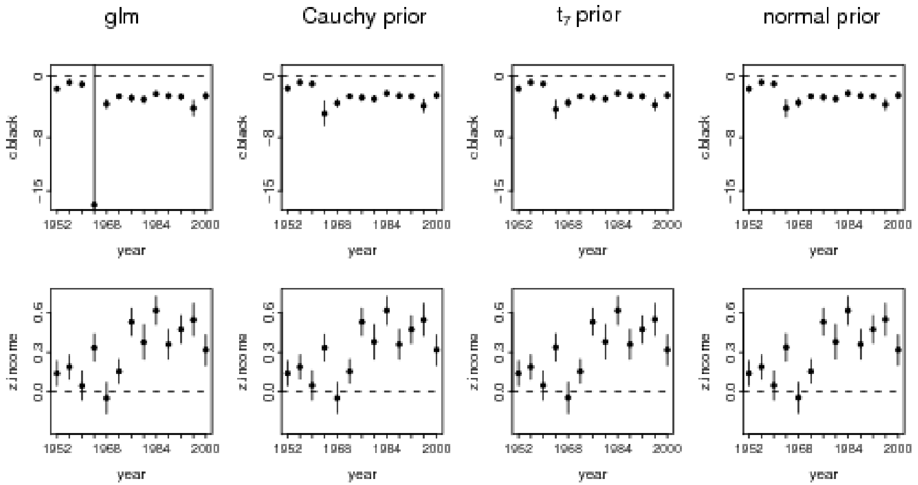
Separation is no joke!

```
glm (vote ~ female + black + income, family=binomial(link="logit"))

1960                                    1968
            coef.est  coef.se                       coef.est  coef.se
(Intercept) -0.14      0.23             (Intercept)  0.47      0.24
female       0.24      0.14             female      -0.01      0.15
black       -1.03      0.36             black       -3.64      0.59
income       0.03      0.06             income      -0.03      0.07

1964                                    1972
            coef.est  coef.se                       coef.est  coef.se
(Intercept)  -1.15      0.22            (Intercept)  0.67      0.18
female       -0.09      0.14            female      -0.25      0.12
black       -16.83    420.40            black       -2.63      0.27
income        0.19      0.06            income       0.09      0.05
```

Regularization in action!



Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
- Weakly informative prior dist

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
    Goal: valid inference for any θ
- Weakly informative prior dist

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist
  - Purposely include less information than we actually have

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist
  - Purposely include less information than we actually have

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist
  - Purposely include less information than we actually have

Information in prior distributions

- Informative prior dist
  - A full generative model for the data
- Noninformative prior dist
  - Let the data speak
  - Goal: valid inference for any θ
- Weakly informative prior dist
  - Purposely include less information than we actually have

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

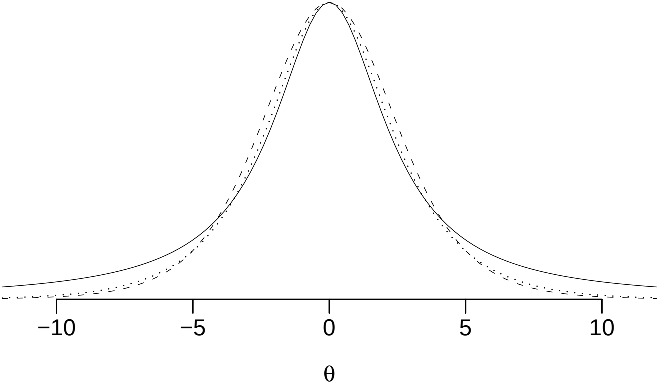
Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`

Prior distributions



Another example

Dose	#deaths/#animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example

Dose	#deaths/#animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example

Dose	#deaths/#animals
-0.86	$0/5$
-0.30	$1/5$
-0.05	$3/5$
0.73	$5/5$

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example

Dose	#deaths/#animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example

Dose	#deaths/#animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

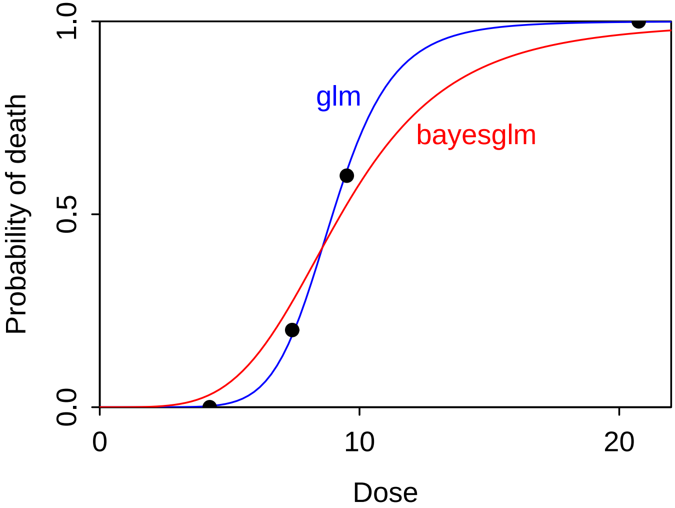
- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example

Dose	#deaths/#animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Maximum likelihood and Bayesian estimates



Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β's is (approx) Cauchy $(0, 1)$
- Our Cauchy $(0, 2.5)$ prior distribution is weakly informative!

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β's is (approx) Cauchy $(0, 1)$
- Our Cauchy $(0, 2.5)$ prior distribution is weakly informative!

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β's is (approx) Cauchy $(0, 1)$
- Our Cauchy $(0, 2.5)$ prior distribution is weakly informative!

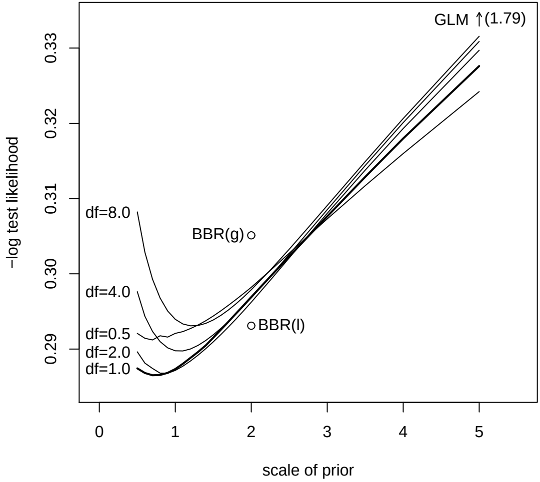
Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β's is (approx) Cauchy $(0, 1)$
- Our Cauchy $(0, 2.5)$ prior distribution is weakly informative!

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β's is (approx) Cauchy $(0, 1)$
- Our Cauchy $(0, 2.5)$ prior distribution is weakly informative!

Expected predictive loss, avg over a corpus of datasets



Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models

No-interaction model

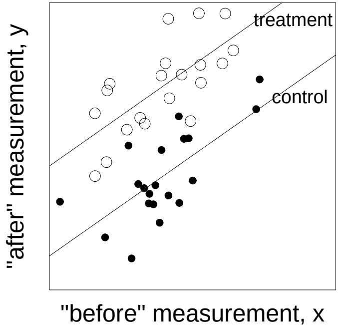
- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$

No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$

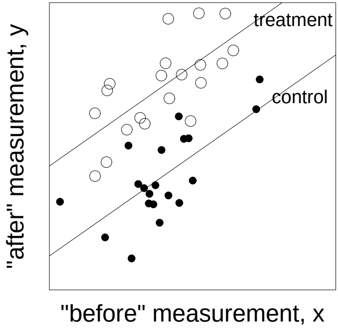
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



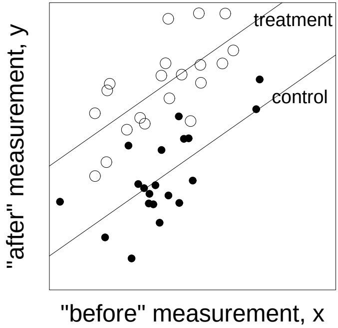
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data

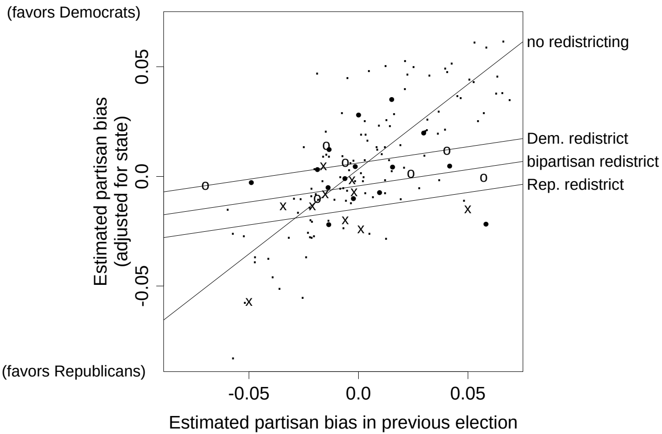
Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data

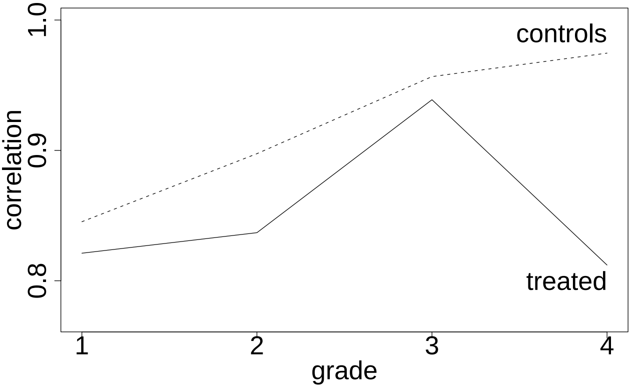
Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data

Observational study of legislative redistricting: before-after data



Educational experiment: correlation between pre-test and post-test data for controls and for treated units



Interactions in regression

- Interactions are important
- Example of income and voting within states (5×50)
- More complicated questions need more elaborate models ($7 \times 5 \times 50$, $2 \times 5 \times 7 \times 50$, ...)

Interactions in regression

- Interactions are important
- Example of income and voting within states (5×50)
- More complicated questions need more elaborate models ($7 \times 5 \times 50$, $2 \times 5 \times 7 \times 50$, $\ldots$)

Interactions in regression

- Interactions are important
- Example of income and voting within states (5×50)
- More complicated questions need more elaborate models ($7 \times 5 \times 50$, $2 \times 5 \times 7 \times 50$, ...)

Interactions in regression

- Interactions are important
- Example of income and voting within states (5×50)
- More complicated questions need more elaborate models ($7 \times 5 \times 50$, $2 \times 5 \times 7 \times 50$, $\dots$)

Red state, blue state, rich state, poor state

- Richer *voters* favor the Republicans, *but*
- Richer *states* favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")

Red state, blue state, rich state, poor state

- Richer *voters* favor the Republicans, *but*
- Richer *states* favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")

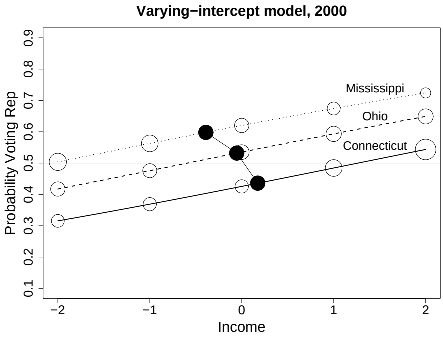
Red state, blue state, rich state, poor state

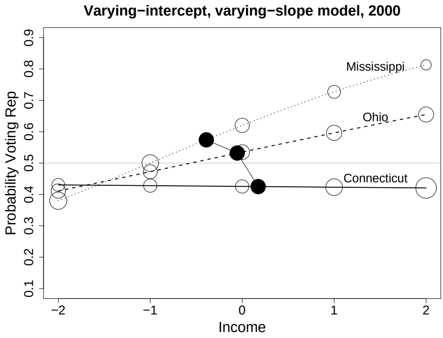
- Richer *voters* favor the Republicans, *but*
- Richer *states* favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")

Red state, blue state, rich state, poor state

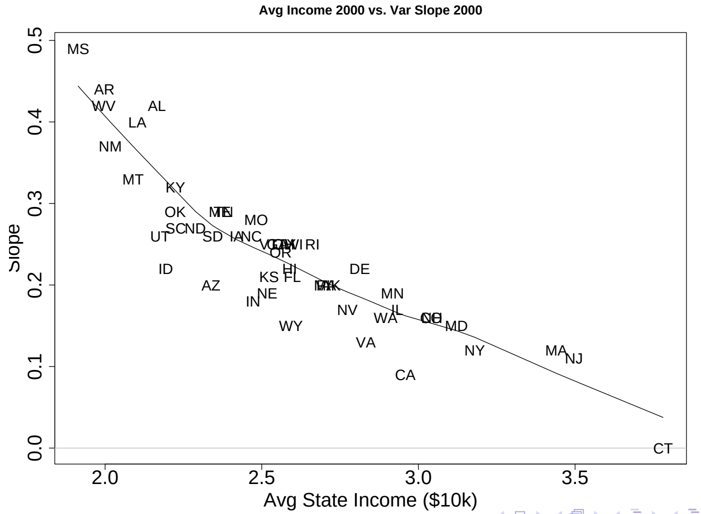
- Richer *voters* favor the Republicans, *but*
- Richer *states* favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")

Varying-intercept model, then model criciticm, then varying-slope model

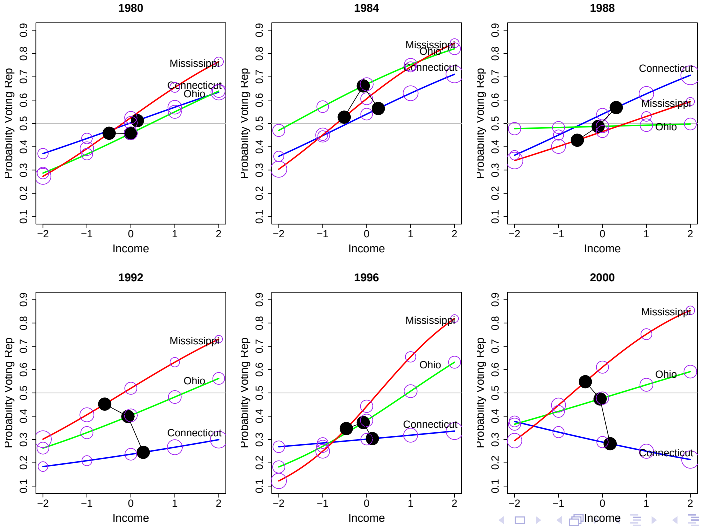




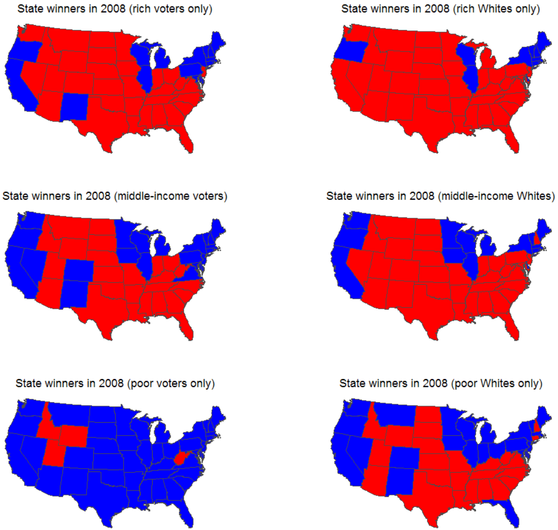
Interactions!



3-way interactions!



Adding another factor: The inference . . .



. . . and the refutation!

- Criticisms from the blogger "Daily Kos":
  - Criticisms of the inferences:
    "While Gelman claims only the under-$20K white demo went for Obama, the results were far different. Per the exit poll – real voters – Obama won all whites: 54-45 percent for those making under $50K, and 51-47% for those making over$50K. . . . New Hampshire is solidly Blue unlike Gelman's maps, 58-40 – one of the most obvious misses in Gelman's analysis. . . ."
  - Criciticms of the method:

. . . and the refutation!

- Criticisms from the blogger "Daily Kos":
  - Criticisms of the inferences:
    "While Gelman claims only the under-$20K white demo went for Obama, the results were far different. Per the exit poll – real voters – Obama won all whites: 54-45 percent for those making under $50K, and 51-47% for those making over$50K. . . . New Hampshire is solidly Blue unlike Gelman's maps, 58-40 – one of the most obvious misses in Gelman's analysis. . . ."
  - Criciticms of the method:
    "Gelman inexplicably avoids using exit poll data . . . while exit polls have their own margin of errors and sample composition problems, they sure as heck beat anything done over the telephone."
  - Traditional statistical "conservatism" will be no defense here!

. . . and the refutation!

- Criticisms from the blogger "Daily Kos":
  - Criticisms of the inferences:
    "While Gelman claims only the under-$20K white demo went for Obama, the results were far different. Per the exit poll – real voters – Obama won all whites: 54-45 percent for those making under $50K, and 51-47% for those making over$50K. . . . New Hampshire is solidly Blue unlike Gelman's maps, 58-40 – one of the most obvious misses in Gelman's analysis. . . ."
  - Criciticms of the method:
    "Gelman inexplicably avoids using exit poll data . . . while exit polls have their own margin of errors and sample composition problems, they sure as heck beat anything done over the telephone."
  - Traditional statistical "conservatism" will be no defense here!

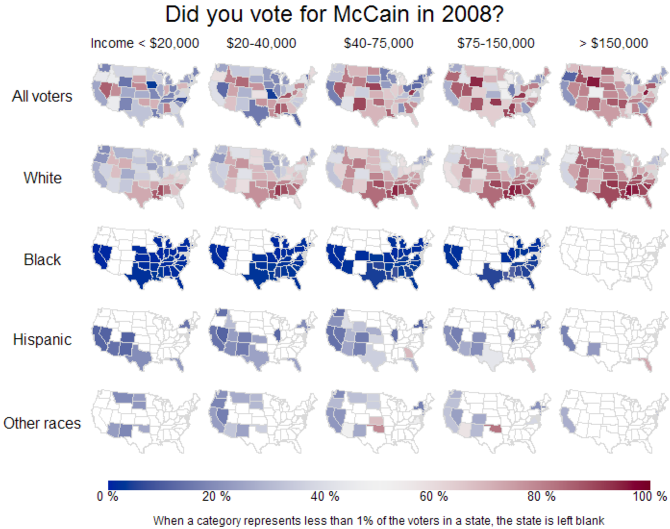
. . . and the refutation!

- Criticisms from the blogger "Daily Kos":
  - Criticisms of the inferences:
    "While Gelman claims only the under-$20K white demo went for Obama, the results were far different. Per the exit poll – real voters – Obama won all whites: 54-45 percent for those making under $50K, and 51-47% for those making over$50K. . . . New Hampshire is solidly Blue unlike Gelman's maps, 58-40 – one of the most obvious misses in Gelman's analysis. . . ."
  - Criciticms of the method:
    "Gelman inexplicably avoids using exit poll data . . . while exit polls have their own margin of errors and sample composition problems, they sure as heck beat anything done over the telephone."
  - Traditional statistical "conservatism" will be no defense here!

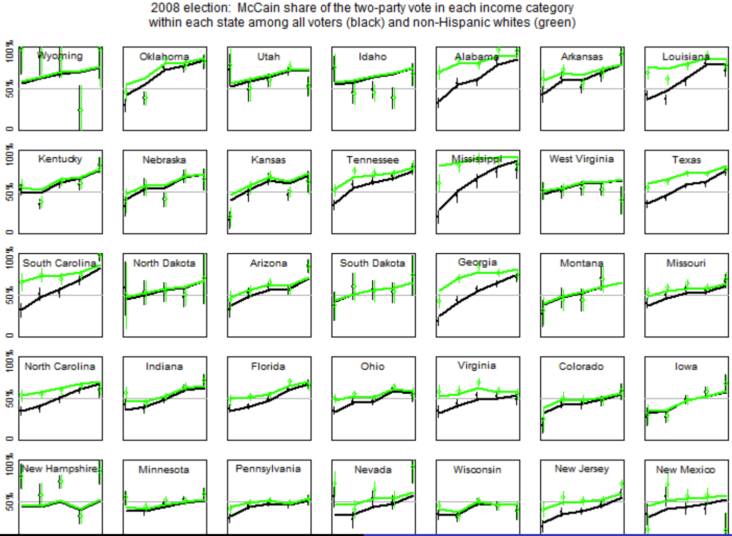
. . . and the refutation!

- Criticisms from the blogger "Daily Kos":
  - Criticisms of the inferences:
    "While Gelman claims only the under-$20K white demo went for Obama, the results were far different. Per the exit poll – real voters – Obama won all whites: 54-45 percent for those making under $50K, and 51-47% for those making over$50K. . . . New Hampshire is solidly Blue unlike Gelman's maps, 58-40 – one of the most obvious misses in Gelman's analysis. . . ."
  - Criciticms of the method:
    "Gelman inexplicably avoids using exit poll data . . . while exit polls have their own margin of errors and sample composition problems, they sure as heck beat anything done over the telephone."
  - Traditional statistical "conservatism" will be no defense here!

After improving the model



A graph we made to study and criticize our inferences



Two more examples

- Ethnicity/religion, income, and school vouchers
  - Show off our method by comparing to (ugly) raw data
- Age, income, and health care

Two more examples

- Ethnicity/religion, income, and school vouchers
  - Show off our method by comparing to (ugly) raw data
- Age, income, and health care
  - Compare to similar graphs of partisanship

Two more examples

- Ethnicity/religion, income, and school vouchers
  - Show off our method by comparing to (ugly) raw data
- Age, income, and health care
  - Compare to similar graphs of partisanship

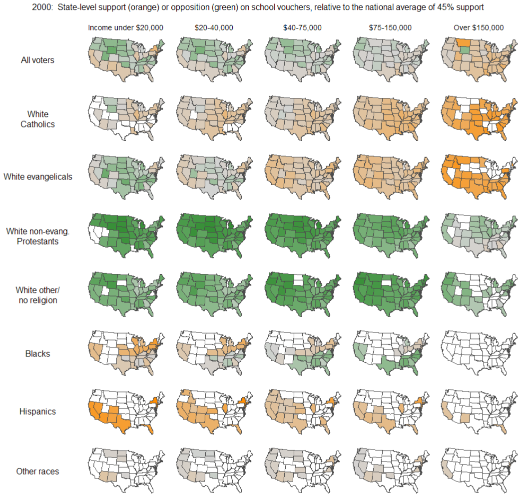
Two more examples

- Ethnicity/religion, income, and school vouchers
  - Show off our method by comparing to (ugly) raw data
- Age, income, and health care
  - Compare to similar graphs of partisanship

Two more examples

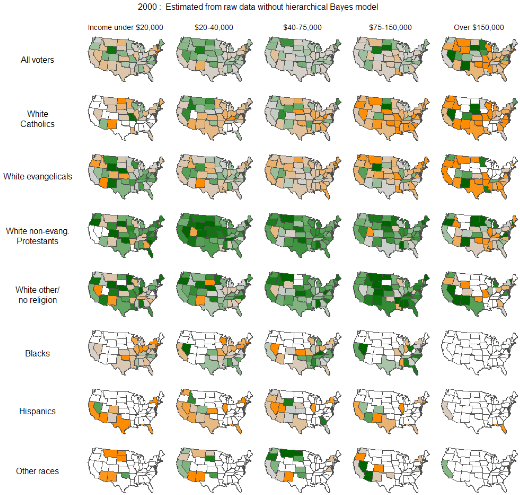
- Ethnicity/religion, income, and school vouchers
  - Show off our method by comparing to (ugly) raw data
- Age, income, and health care
  - Compare to similar graphs of partisanship

Ethnicity/religion, income, and school vouchers



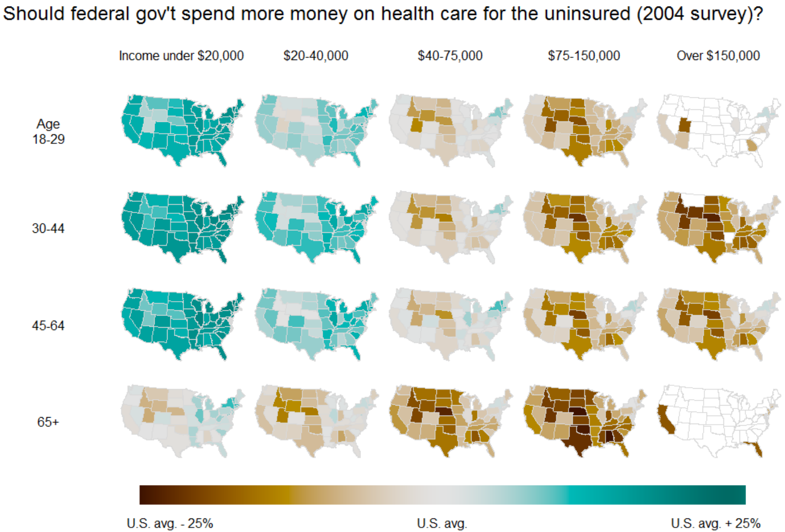
Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnic/religious cagetories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters of a state, the state is left blank.

The raw data



Compared to the Bayes maps, these are very noisy, and it is difficult to try to interpret the patterns.

Age, income, and health care



Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter *matrices* α_{jk} don't look like exchangeable *vectors*
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, ...
- Recall the "blessing of dimensionality": as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model

Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter *matrices* α_{jk} don't look like exchangeable *vectors*
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, . . .
- Recall the "blessing of dimensionality": as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model

Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter *matrices* α_{jk} don't look like exchangeable *vectors*
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, ...
- Recall the "blessing of dimensionality": as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model

Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter *matrices* α_{jk} don't look like exchangeable *vectors*
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, ...
- Recall the "blessing of dimensionality": as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model

Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter *matrices* α_{jk} don't look like exchangeable *vectors*
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, . . .
- Recall the "blessing of dimensionality": as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

What have we learned?

- Models need structure but not too much structure
- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Conservatism in statistics
- Weak prior information is key

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
    - Theoretical statisticians
    - Applied statisticians
    - Computer scientists
- Same data structure, different models
    - Physics
    - Political science
    - Economics
    - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health

Cultural differences

- How do you motivate/justify/defend/promote a statistical method?
  - Theoretical statisticians
  - Applied statisticians
  - Computer scientists
- Same data structure, different models
  - Physics
  - Political science
  - Economics
  - Biology/health