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

\[ y = \logit^{-1}(x) \]

slope = 1/4
A clean example

estimated $\Pr(y=1) = \logit^{-1}(-1.40 + 0.33 \times)$

slope = 0.33/4
The problem of separation

\[
\text{slope} = \infty? \quad \text{graph with data points}
\]
Separation is no joke!

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

<table>
<thead>
<tr>
<th></th>
<th>1960 coef.est</th>
<th>1960 coef.se</th>
<th>1968 coef.est</th>
<th>1968 coef.se</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.14</td>
<td>0.23</td>
<td>0.47</td>
<td>0.24</td>
</tr>
<tr>
<td>female</td>
<td>0.24</td>
<td>0.14</td>
<td>-0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>black</td>
<td>-1.03</td>
<td>0.36</td>
<td>-3.64</td>
<td>0.59</td>
</tr>
<tr>
<td>income</td>
<td>0.03</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1964 coef.est</th>
<th>1964 coef.se</th>
<th>1972 coef.est</th>
<th>1972 coef.se</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.15</td>
<td>0.22</td>
<td>0.67</td>
<td>0.18</td>
</tr>
<tr>
<td>female</td>
<td>-0.09</td>
<td>0.14</td>
<td>-0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>black</td>
<td>-16.83</td>
<td>420.40</td>
<td>-2.63</td>
<td>0.27</td>
</tr>
<tr>
<td>income</td>
<td>0.19</td>
<td>0.06</td>
<td>0.09</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Regularization in action!

Andrew Gelman
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 $\theta$
- 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 $\theta$
- 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 $\theta$
- 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 $\theta$
- 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 $\theta$
- 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 $\theta$
- 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 $\theta$
- 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 coefficients 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 distributions with center 0 and scale 2.5
  - Rescale each predictor to have mean 0 and standard deviation $\frac{1}{2}$
  - Fast implementation using EM; easy adaptation of glm
Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefficients 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 distributions with center $0$ and scale $2.5$
- Rescale each predictor to have mean $0$ and standard deviation $\frac{1}{2}$
- 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 coeffs 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 coefficients 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 distributions with center $0$ and scale $2.5$
  - Rescale each predictor to have mean $0$ and standard deviation $\frac{1}{2}$
  - Fast implementation using EM; easy adaptation of glm
Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefficients 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 distributions 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

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is $7.8 \pm 4.9$
  - With weakly-informative prior: Bayes est is $4.4 \pm 1.9$

Which is truly conservative?

The sociology of shrinkage
Another example

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

- Slope of a logistic regression of \( \text{Pr(death)} \) on dose:
  - Maximum likelihood est is \( 7.8 \pm 4.9 \)
  - With weakly-informative prior: Bayes est is \( 4.4 \pm 1.9 \)

- Which is truly conservative?
- The sociology of shrinkage
Another example

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is $7.8 \pm 4.9$
  - With weakly-informative prior: Bayes est is $4.4 \pm 1.9$

- Which is truly conservative?
- The sociology of shrinkage
Another example

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Dose</th>
<th>#deaths/#animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.86</td>
<td>0/5</td>
</tr>
<tr>
<td>−0.30</td>
<td>1/5</td>
</tr>
<tr>
<td>−0.05</td>
<td>3/5</td>
</tr>
<tr>
<td>0.73</td>
<td>5/5</td>
</tr>
</tbody>
</table>

- Slope of a logistic regression of $\Pr(\text{death})$ on dose:
  - Maximum likelihood est is $7.8 \pm 4.9$
  - With weakly-informative prior: Bayes est is $4.4 \pm 1.9$
- Which is truly conservative?
- The sociology of shrinkage
Maximum likelihood and Bayesian estimates

![Graph showing maximum likelihood (glm) and Bayesian estimates (bayesglm) for the probability of death as a function of dose.]

- **Dose**: 0, 10, 20
- **Probability of death**: 0.0, 0.5, 1.0

Andrew Gelman
Creating structured and flexible models: some open problems
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 $\beta$'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 $\beta$’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 $\beta$’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 $\beta$’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 $\beta$’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

![Graph showing expected predictive loss](image-url)
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: $y_i = T_i \theta + X_i \beta + \epsilon_i$

```
"before" measurement, x

"after" measurement, y
```

Andrew Gelman

Creating structured and flexible models: some open problems
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: $y_i = T_i \theta + X_i \beta + \epsilon_i$

![Graph showing before-after measurements with control and treatment groups.](image)
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

Andrew Gelman
Creating structured and flexible models: some open problems
Interactions in regression

- Interactions are important
- Example of income and voting within states ($5 \times 50$)
- 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 ($5 \times 50$)
- 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 × 5 × 50, 2 × 5 × 7 × 50, ...)

Andrew Gelman
Creating structured and flexible models: some open problems
Interactions in regression

- Interactions are important
- Example of income and voting within states (5 × 50)
- More complicated questions need more elaborate models (7 × 5 × 50, 2 × 5 × 7 × 50, . . .)
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 criticism, then varying-slope model

Varying-intercept model, 2000

Varying-intercept, varying-slope model, 2000
Interactions!

Avg Income 2000 vs. Var Slope 2000

Avg State Income ($10k) vs. Slope
3-way interactions!

Income and voting
Ethnicity/religion, income, and school vouchers
Age, income, and health care

Andrew Gelman
Creating structured and flexible models: some open problems
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. ...”
  - Criticisms 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!

Andrew Gelman
Creating structured and flexible models: some open problems
... 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. . . .”
  - Criticisms 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. ...”
  - Criticisms 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. ...”
  - Criticisms 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!
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. . . . ”

- Criticisms 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

Did you vote for McCain in 2008?

Income < $20,000  $20-40,000  $40-75,000  $75-150,000  > $150,000

All voters

White

Black

Hispanic

Other races

When a category represents less than 1% of the voters in a state, the state is left blank
A graph we made to study and criticize our inferences

2008 election: McCain share of the two-party vote in each income category within each state among all voters (black) and non-Hispanic whites (green)
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
The raw data

2000 - Estimated from raw data without hierarchical Bayes model

- Income under $20,000
- $20-40,000
- $40-75,000
- $75-150,000
- Over $150,000

- All voters
- White Catholics
- White evangelicals
- White non-evang. Protestants
- White other/no religion
- Blacks
- Hispanics
- Other races

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

Should federal gov't spend more money on health care for the uninsured (2004 survey)?

Income under $20,000 | $20-40,000 | $40-75,000 | $75-150,000 | Over $150,000

Age 18-29

Age 30-44

Age 45-64

Age 65+

Andrew Gelman

Creating structured and flexible models: some open problems
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices \( \alpha_{jk} \) 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 $\alpha_{jk}$ 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 $\alpha_{jk}$ 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 $\alpha_{jk}$ 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 $\alpha_{jk}$ 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