Dealing with Separation in Logistic Regression Models

Carlisle Rainey
Assistant Professor
Texas A&M University
crainey@tamu.edu

paper, data, and code at
crain.co/research
The prior matters a lot, so choose a good one.
The prior matters a lot,

1. in practice
2. in theory

so choose a good one.

3. concepts
4. software
The Prior Matters in Practice
politics need
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Governor</td>
<td>-26.35</td>
<td>[-126,979.03; 126,926.33]</td>
</tr>
<tr>
<td>% Uninsured (Std.)</td>
<td>0.92</td>
<td>[-3.46; 5.30]</td>
</tr>
<tr>
<td>% Favorable to ACA</td>
<td>0.01</td>
<td>[-0.17; 0.18]</td>
</tr>
<tr>
<td>GOP Legislature</td>
<td>2.43</td>
<td>[-0.47; 5.33]</td>
</tr>
<tr>
<td>Fiscal Health</td>
<td>0.00</td>
<td>[-0.02; 0.02]</td>
</tr>
<tr>
<td>Medicaid Multiplier</td>
<td>-0.32</td>
<td>[-2.45; 1.80]</td>
</tr>
<tr>
<td>% Non-white</td>
<td>0.05</td>
<td>[-0.12; 0.21]</td>
</tr>
<tr>
<td>% Metropolitan</td>
<td>-0.08</td>
<td>[-0.17; 0.02]</td>
</tr>
<tr>
<td>Constant</td>
<td>2.58</td>
<td>[-7.02; 12.18]</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Democratic Governor</td>
<td>-26.35</td>
<td>[-126,979.03; 126,926.33]</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Democratic Governor</td>
<td>-26.35</td>
<td>[-126,979.03; 126,926.33]</td>
</tr>
</tbody>
</table>

This is a failure of maximum likelihood.
Those inferences from supposedly "default" priors aren't even close to similar. Very sad!
Different default priors produce different results.
The Prior Matters in Theory
For

1. a monotonic likelihood \( p(y|\beta) \) decreasing in \( \beta_s \),
2. a proper prior distribution \( p(\beta|\sigma) \), and
3. a large, negative \( \beta_s \),

the posterior distribution of \( \beta_s \) is proportional to the prior distribution for \( \beta_s \), so that \( p(\beta_s|y) \propto p(\beta_s|\sigma) \).
For

1. a monotonic likelihood \( p(y|\beta) \) decreasing in \( \beta_s \),
2. a proper prior distribution \( p(\beta|\sigma) \), and
3. a large, negative \( \beta_s \),

the posterior distribution of \( \beta_s \) is proportional to the prior distribution for \( \beta_s \), so that \( p(\beta_s|y) \propto p(\beta_s|\sigma) \).
The prior determines crucial parts of the posterior.
Key Concepts for Choosing a Good Prior
Pr(y_i) = \Lambda(\beta_c + \beta_s s_i + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})
Transforming the Prior Distribution

\[ \tilde{\beta} \sim p(\beta) \]

\[ \tilde{\pi}_{new} = p(y_{new} | \tilde{\beta}) \]

\[ \tilde{q}_{new} = q(\tilde{\pi}_{new}) \]
We Already Know Few Things

\[ \beta_1 \approx \hat{\beta}_{1}^{mle} \]
\[ \beta_2 \approx \hat{\beta}_{2}^{mle} \]
\[ \vdots \]
\[ \beta_k \approx \hat{\beta}_{k}^{mle} \]

\[ \beta_s < 0 \]
Partial Prior Distribution

\[ p^*(\beta | \beta_s < 0, \beta_{-s} = \hat{\beta}_{mle}^{s}) \],

where \( \hat{\beta}_{mle}^s = -\infty \)
Software for Choosing a Good Prior
separation
(on GitHub)
Stan Project

rstanarm

StataStan
Conclusion
The prior matters a lot, so choose a good one.
What should you do?

1. Notice the problem and do something.

2. Recognize the prior affects the inferences and choose a good one.

3. Assess the robustness of your conclusions to a range of prior distributions.