Approximate Bayesian logistic regression via penalized likelihood estimation with data augmentation

Andrea Discacciati   Nicola Orsini

Unit of Biostatistics and Unit of Nutritional Epidemiology
Institute of Environmental Medicine
Karolinska Institutet
http://www.imm.ki.se/biostatistics/
andrea.discacciati@ki.se

2014 Italian Stata Users Group meeting

13th November 2014
Background

- Bayesian analyses are uncommon in epidemiological research
- Partly because of the absence of Bayesian methods from most basic courses in statistics...
- ...but also because of the misconception that they are computationally difficult and require specialized software
- However, approximate Bayesian analyses can be carried out using standard software for frequentist analyses (e.g.: Stata)
- This can be done through penalized likelihood estimation, which in turn can be implemented via data augmentation
Aims of this presentation

- Introduce penalized likelihood (PL) estimation in the context of logistic regression
- Present a new Stata command (`penlogit`) that fits penalized logistic regression via data augmentation
- Show a practical example of a Bayesian analysis using `penlogit`
How to fit a Bayesian model

A partial list (in order of increasing “exactness”):

- Monte Carlo sensitivity analysis
- Inverse-variance weighting (information-weighted averaging)
- Penalized likelihood
- Posterior sampling (e.g.: Markov chain Monte Carlo (MCMC))
Penalized log-likelihood

- A penalized log-likelihood (PLL) is a log-likelihood with a penalty function added to it

**PLL for a logistic regression model**

\[
\ln [L(\beta; x)] + P(\beta) = \sum_i \left\{ \ln [\text{expit}(x_i^T \beta)] y_i + \ln [1 - \text{expit}(x_i^T \beta)] (n_i - y_i) \right\} + P(\beta)
\]

- \( \beta = \{\beta_1, \ldots, \beta_p\} \) is the vector of unknown regression coefficients
- \( \ln (L(\beta; x)) \) is the log-likelihood of a standard logistic regression
- \( P(\beta) \) is the penalty term
- The penalty \( P(\beta) \) pulls or shrinks the final estimates away from the ML estimates, toward \( m = \{m_1, \ldots, m_p\} \)
Bayesian perspective

Link between PLL and Bayesian framework

We add the logarithm of the prior density function $f(\beta)$ as the penalty term $P(\beta)$ in the log-likelihood

- A prior for a parameter $\beta_i$ is a probability distribution that reflects one’s uncertainty about $\beta_i$ before the data under analysis is taken into account
- Two extreme cases: priors with $+\infty$ variance and priors with 0 variance
Normal priors

- Normal priors for $\beta_i$ (ln(OR)): $\beta_i \sim N(m_i, v_i)$
- These priors are symmetric and unimodal
- $m_i=$ mean=median=mode
- Amount of background information controlled by the variance $v_i$
- Equivalently, these are log-normal priors on the OR scale ($\exp(\beta_i)$)

Penalty function

$$P(\tilde{\beta}) = -\frac{1}{2} \left[ \sum_{j=1}^{q} \frac{1}{\nu_j} (\beta_j - m_j)^2 \right]$$
Generalized log-F priors

- Characterized by 4 parameters: $\beta_i \sim \log-F(m_i, df_{1,i}, df_{2,i}, s_i)$
- These priors are unimodal ($m_i$), but can be skewed (increasing the difference between $df_{1,i}$ and $df_{2,i}$)
- Log-F priors are more flexible than normal priors and are useful for example when prior information is directional
Posterior distribution and PLL

The PLL is, apart from an additive constant, equal to the logarithm of the posterior distribution of $\beta$ given the data.

- In terms of PL: $PL(\beta; x) \propto f(\beta|x) = k \times L(\beta; x) \times \prod_j f_j(\beta_j)$
- Maximum PL estimate of $\beta$ ($\beta_{post}$) is the maximum a posteriori estimate
- $100(1 - \alpha)\%$ Wald CL are the approximate posterior limits, i.e. the $\frac{\alpha}{2}$ and $\left(1 - \frac{\alpha}{2}\right)$ quantiles of the posterior distribution
- If the profile PLL of $\beta_i$ is not closely quadratic, it is better to use penalized profile-likelihood limits to approximate posterior limits
Data-augmentation priors (DAPs)

- Algebraically equivalent way of maximizing the PLL is using DAPs
- Prior distributions on the parameters are represented by prior data records created ad hoc
- Prior data records generate a penalty function that imposes the desired priors on the model parameters
- Estimation carried out using standard ML machinery on the augmented dataset (i.e. original and DAP records)

Advantage of PL estimation via DAPs

By translating prior distributions to equivalent data, DAPs are one way of understanding the logical strength of the imposed priors
penlogit — a brief overview

Description

penlogit provides estimates for the penalized logistic model, whose PLL was defined in slide 5, using data augmentation priors

- Specify a binary outcome and one or more covariates
- Priors can be imposed using the nprior and lfprior options
- Penalized profile-likelihood limits can be obtained with the ppl option
- net install penlogit,
  from(http://www.imm.ki.se/biostatistics/stata/)
The data

- Data from a study of obstetric care and neonatal death \( (n = 2992) \)
- The full dataset includes a total of 14 covariates
- Univariate analysis: hydramnios during pregnancy as the exposure

<table>
<thead>
<tr>
<th></th>
<th>Hydramnios</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( X = 1 )</td>
<td>( X = 0 )</td>
<td>Total</td>
</tr>
<tr>
<td>Deaths ( (Y = 1) )</td>
<td>1</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Survivals ( (Y = 0))</td>
<td>9</td>
<td>2,966</td>
<td>2,975</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>2,982</td>
<td>2,992</td>
</tr>
</tbody>
</table>

- Sparse data (only one exposed case)
Frequentist analysis

- No explicit prior on $\beta_{hydram}$
- This corresponds to an implicit prior $N(0, +\infty)$
- This prior gives equal odds on $OR = 10^{-100}$, $OR = 1$ or $OR = 10^{100}$

Logistic regression

| death | Coef.   | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------|---------|-----------|------|-------|----------------------|
| hydram | 3.025156 | 1.083489  | 2.79 | 0.005 | .9015571 5.148755   |

<table>
<thead>
<tr>
<th>death</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>[95% PLL Conf. Int.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydram</td>
<td>3.025156</td>
<td>1.199495</td>
<td>.0819808 4.783916</td>
</tr>
</tbody>
</table>

- $OR = 20.6$ (95% profile-likelihood C.I.: 1.08, 119)
- Profile-likelihood function for $\beta_{hydram}$ is strongly asymmetrical
Specifying the prior for $\beta_{\text{hydram}}$

- Normal prior on $\beta_{\text{hydram}}$
- Prior information was expressed in terms of 95% prior limits on the OR scale: (1, 16)
- Under normality, it is easy to calculate the corresponding hyperparameters $m_{\text{hydram}}$ and $v_{\text{hydram}}$ that yield those 95% prior limits
  - $\beta_{\text{hydram}} \sim N(\ln(4), 0.5)$
- Semi-Bayes analysis because we do not impose a prior on the intercept $\beta_0$
Direct PLL maximization

PLL maximized using `mlexp` in Stata 13

```
mlexp (log(invlogit({b0}+{xb:hydram}))*death +
    log(1-(invlogit({b0}+{xb:}))*(1-death) -
    0.5*0.5^(-1)*(xb_hydram-log(4))^2/2992))
```

```
. lincom [xb_hydram]_cons, or

| Odds Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|------------|-----------|------|------|----------------------|
| (1)        | 5.652566  | 3.732409 | 2.62 | 0.009 | 1.54951 20.62039 |
```

- $\text{OR}_{post}$ (95% Wald posterior limits) $= 5.65$ (1.55, 20.6)

Andrea Discacciati

Karolinska Institutet
PLL estimation via DAPs

- Data augmentation has the advantage of showing the strength of the prior being imposed.
- It shows the number of cases and noncases that would supply data information about the coefficient approximately equivalent to the information supplied by the prior.
- The prior $N(\ln(4), 0.5)$ supplies data information roughly equivalent to 4.5 cases and 4.5 noncases (see penlogit output in the next slide).
**PLL estimation via DAPs**

**penlogit Stata command**

```stata
penlogit death hydram, nprior(hydram ln(4) 0.5) ppl(hydram) or
```

Penalized logistic regression

```
Normal prior for hydram: exact prior median OR (95% PL): 4.00 (1.00, 16.00)
Data approx. equivalent to prior: cases=4.54 noncases=4.54 exp(offset)=.912
```

|          | Odds Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|------------|-----------|-------|-----|----------------------|
| death    |            |           |       |     |                      |
| hydram   | 5.652642   | 3.732672  | 2.62  | 0.009 | 1.549416 – 20.6222   |

<table>
<thead>
<tr>
<th></th>
<th>[95% PL Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>death</td>
<td></td>
</tr>
<tr>
<td>hydram</td>
<td>1.509324 – 19.84511</td>
</tr>
</tbody>
</table>

- \( OR_{post} \) (95% PL posterior limits) = 5.65 (1.50, 19.8)
- Similar to the Wald posterior limits because of the symmetrizing effect of the normal prior
MCMC and comparison of the results

- MCMC analysis carried out using OpenBUGS called from within Stata (see John Thompson’s commands: \texttt{wbsrun}, \texttt{wbsscript}, \ldots)
- 1 chain, 20,000 samples form the posterior distribution
- Results, not surprisingly, are virtually identical

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Approximate posterior percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50th</td>
</tr>
<tr>
<td>Direct PLE (\texttt{mlexp})†</td>
<td>5.652</td>
</tr>
<tr>
<td>PLE via DAPs (\texttt{penlogit})‡</td>
<td>5.652</td>
</tr>
<tr>
<td>MCMC (OpenBUGS)</td>
<td>5.595</td>
</tr>
</tbody>
</table>

†: 95% Wald posterior limits  ‡: 95% penalized profile-likelihood posterior limits
Multivariate analysis: specifying the priors

- 14 covariates
- The model parameters were given three possible priors
- They reflected the background clinical information on the different risk factors of neonatal death

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Variable name</th>
<th>Prior</th>
<th>Prior percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past abortion</td>
<td>abort</td>
<td>Normal(0,0.5)</td>
<td>1.00 0.25 4.00</td>
</tr>
<tr>
<td>No monitor</td>
<td>nomonit</td>
<td>Normal(ln(2),0.5)</td>
<td>2.00 0.50 8.00</td>
</tr>
<tr>
<td>Early age</td>
<td>teenagers</td>
<td>Normal(ln(2),0.5)</td>
<td>2.00 0.50 8.00</td>
</tr>
<tr>
<td>Hydramnios</td>
<td>hydram</td>
<td>Normal(ln(4),0.5)</td>
<td>4.00 1.00 16.00</td>
</tr>
<tr>
<td>Twin, triplet</td>
<td>twint</td>
<td>Normal(ln(4),0.5)</td>
<td>4.00 1.00 16.00</td>
</tr>
</tbody>
</table>
PLL estimation via DAPs

- With `penlogit` it is easy to specify the priors on the 14 coefficients

```
penlogit death abort nomonit teenagers [...] hydram twint,
nprior(abort 0 0.5 nomonit ln(2) 0.5 teenagers ln(2) 0.5
[...] hydram ln(4) 0.5 twint ln(4) 0.5)
ppl(nomonit teenagers [...] hydram twint) or
```
## Approximate posterior percentiles

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Variable name</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Data augmentation</td>
<td>MCMC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50th</td>
<td>2.5th</td>
<td>97.5th</td>
<td>50th</td>
<td>2.5th</td>
<td>97.5th</td>
</tr>
<tr>
<td>Past abortion</td>
<td>abort</td>
<td>0.83</td>
<td>0.31</td>
<td>1.9</td>
<td>0.79</td>
<td>0.29</td>
<td>1.9</td>
</tr>
<tr>
<td>No monitor</td>
<td>nomonit</td>
<td>1.7</td>
<td>0.68</td>
<td>4.8</td>
<td>1.8</td>
<td>0.71</td>
<td>5.0</td>
</tr>
<tr>
<td>Early age</td>
<td>teenagers</td>
<td>1.6</td>
<td>0.61</td>
<td>4.0</td>
<td>1.6</td>
<td>0.59</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydramnios</td>
<td>hydram</td>
<td>6.1</td>
<td>1.6</td>
<td>23</td>
<td>6.0</td>
<td>1.6</td>
<td>22</td>
</tr>
<tr>
<td>Twin, triplet</td>
<td>twint</td>
<td>5.2</td>
<td>1.8</td>
<td>14</td>
<td>5.3</td>
<td>1.8</td>
<td>14</td>
</tr>
</tbody>
</table>

- Again, posterior percentiles from PLE via DAPs and from MCMC showed exceptionally good agreement
Prior, posterior, and profile-likelihood for $\beta_{hydram}$

- The posterior distribution is almost perfectly symmetric because of the symmetrizing effect of the normal prior.
Strengths of PLE via DAPs for Bayesian analyses

- Does not require the use of specialized software
- Computationally easier than simulation methods (e.g.: MCMC)
- Also useful for Bayesian sensitivity analyses and to provide reasonable starting values and convergence checks for MCMC
- DAPs provide a critical perspective on the proposed priors

Caveats

- Approximate posterior mode and 95% posterior limits (but adequate in the context of observational epidemiology)
- Uses same large-sample approximations as ML (but more stable thanks to the stabilizing and symmetrizing effect of the penalty)
- Profile-posterior limits if the posterior distribution is non-normal
References

- Discacciati, A., Orsini, N., and Greenland, S. Approximate Bayesian logistic regression via penalized likelihood estimation with data augmentation. Submitted to the *Stata Journal*.


