Logistic regression in small samples

(Nearly) Unbiased estimation with penalized maximum likelihood

Kelly McCaskey and Carlisle Rainey
paper at kellymccaskey.com/small
code and data at github.com/kellymccaskey/small

TEXAS A&M UNIVERSITY

The Problem: Increased Bias and Variance

When modeling a binary outcome, a researcher might use logistic regression, and model the probability of an event as

$$\Pr(y_i) = \Pr(y_i = 1 \mid X_i) = \frac{1}{1 + e^{-X_i\beta}} \ ,$$

where y represents a vector of binary outcomes, X a matrix of explanatory variables and an intercept, and β as a vector of model coefficients.

It is straightforward, then, to derive the likelihood function

$$\Pr(y|\beta) = L(\beta|y) = \prod_{i=1}^{n}\left[\left(\frac{1}{1+e^{-X_i\beta}}\right)^{y_i}\left(1 - \frac{1}{1+e^{-X_i\beta}}\right)^{1-y_i}\right].$$

Per usual, one can then take the natural logarithm of both sides to obtain the log-likelihood function

$$\log L(\beta|y) = \sum_{i=1}^{n}\left[y_i \log\left(\frac{1}{1+e^{-X_i\beta}}\right) + (1-y_i)\log\left(1 - \frac{1}{1+e^{-X_i\beta}}\right)\right].$$

The researcher can then obtain the maximum likelihood (ML) estimate $\hat{\beta}^{mle}$ by finding the vector β that maximizes $\log L$ (King 1998).

Asymptotically, ML estimates for logistic regression coefficients are unbiased. For **small samples**, though, the asymptotic approximation does not work well — **the ML estimates are biased substantially away from zero**.

Key Question: How can the researcher obtain unbiased estimates of logistic regression coefficients with small samples?

Firth's Solution: Penalized Maximum Likelihood

Firth (1993) suggests penalizing the usual likelihood function by a factor equal to the square root of the determinant of the information matrix, producing a penalized likelihood function

$$L^*(\beta|y) = L(\beta|y)|I(\beta)|^{\frac{1}{2}}.$$

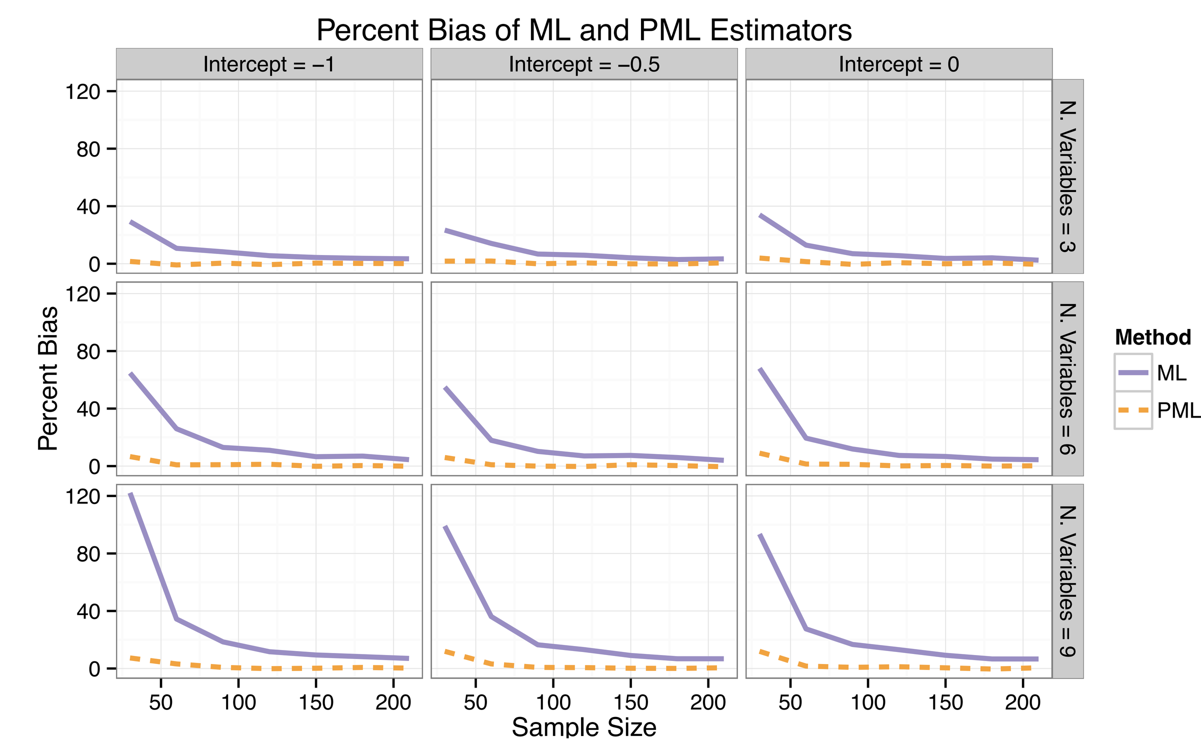
Taking the natural log yields the penalized log-likelihood function

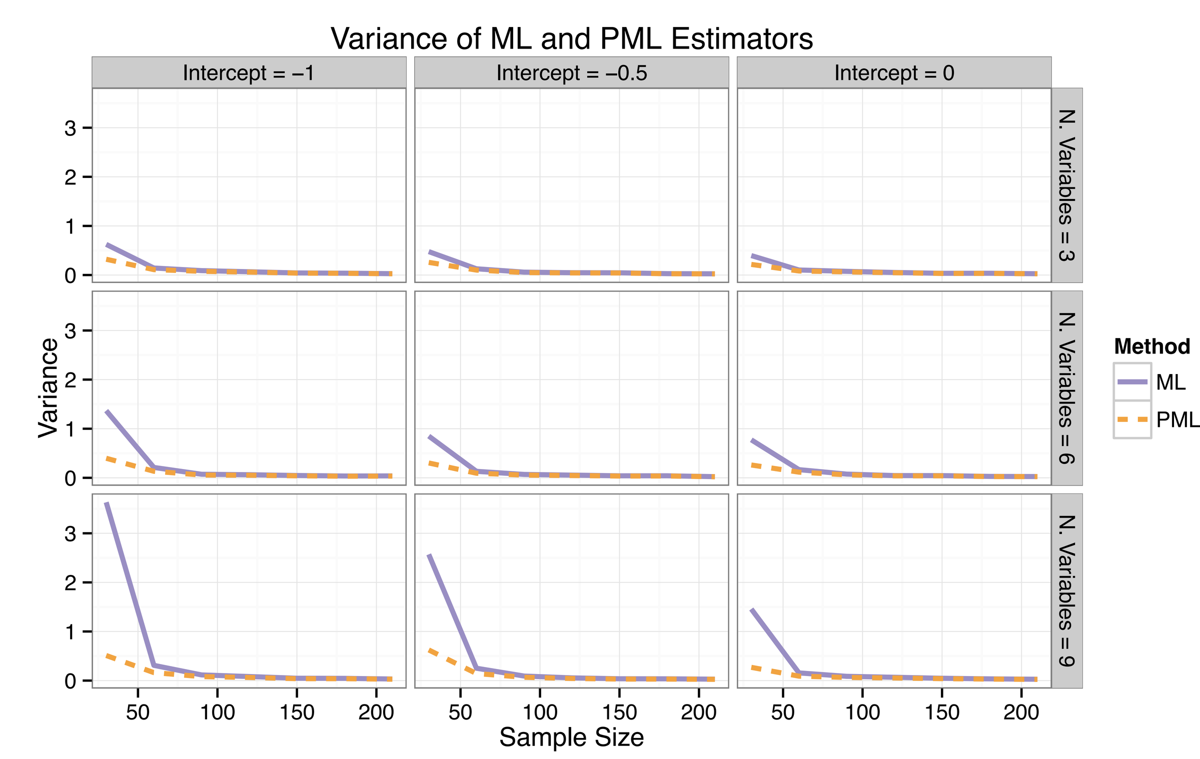
$$\log L^*(\beta|y) = \sum_{i=1}^{n}\left[y_i \log\left(\frac{1}{1+e^{-X_i\beta}}\right) + (1-y_i)\log\left(1 - \frac{1}{1+e^{-X_i\beta}}\right)\right] + \frac{1}{2}\log|I(\beta)|.$$

The researcher can then obtain the *penalized* maximum likelihood (PML) estimate $\hat{\beta}^{pmle}$ by finding the vector β that maximizes $\log L^*$.

Key Point: Using PML, the researcher can obtain nearly unbiased estimates of logistic regression coefficients. The PML estimates also exhibit less variance than the ML estimates..

Monte Carlo Simulations





Data Generating Process (DGP)

The DGP remains constant in our simulation of 10,000 data sets, where:

- $\Pr(y_i = 1) = \frac{1}{1+e^{-X_i\beta}}$, $i \in 1, 2, ..., n$ and $X_i\beta = \beta_{cons} + 0.5x_1 + \sum_{j=2}^{k} 0.2x_j$.
- The coefficient for x_1 is the coefficient of interest.
- Each fixed x_j is drawn from a normal distribution with mean of zero and standard deviation of one.
- Sample size n varies from 30 to 210, the number of explanatory variables k vary from 3 to 6 to 9, and the intercept β_{cons} varies from -1 to -0.5 to 0.

Bias and Variance

1. The ML estimates can sometimes **be biased by more than 50%** while PML remains **nearly unbiased regardless of sample size.**
2. The variance of the ML estimator can be three or more times as large as the PML estimator.

Application: A Reanalysis of Weisiger 2014

Weisiger (2014) describes how, after the official end of the war, violence sometimes continues in the form of guerrilla warfare when conditions are favorable for insurgency.

Hypothesis

Post-war resistance is more likely when the pre-war leader remains at large within the country.

Using PML

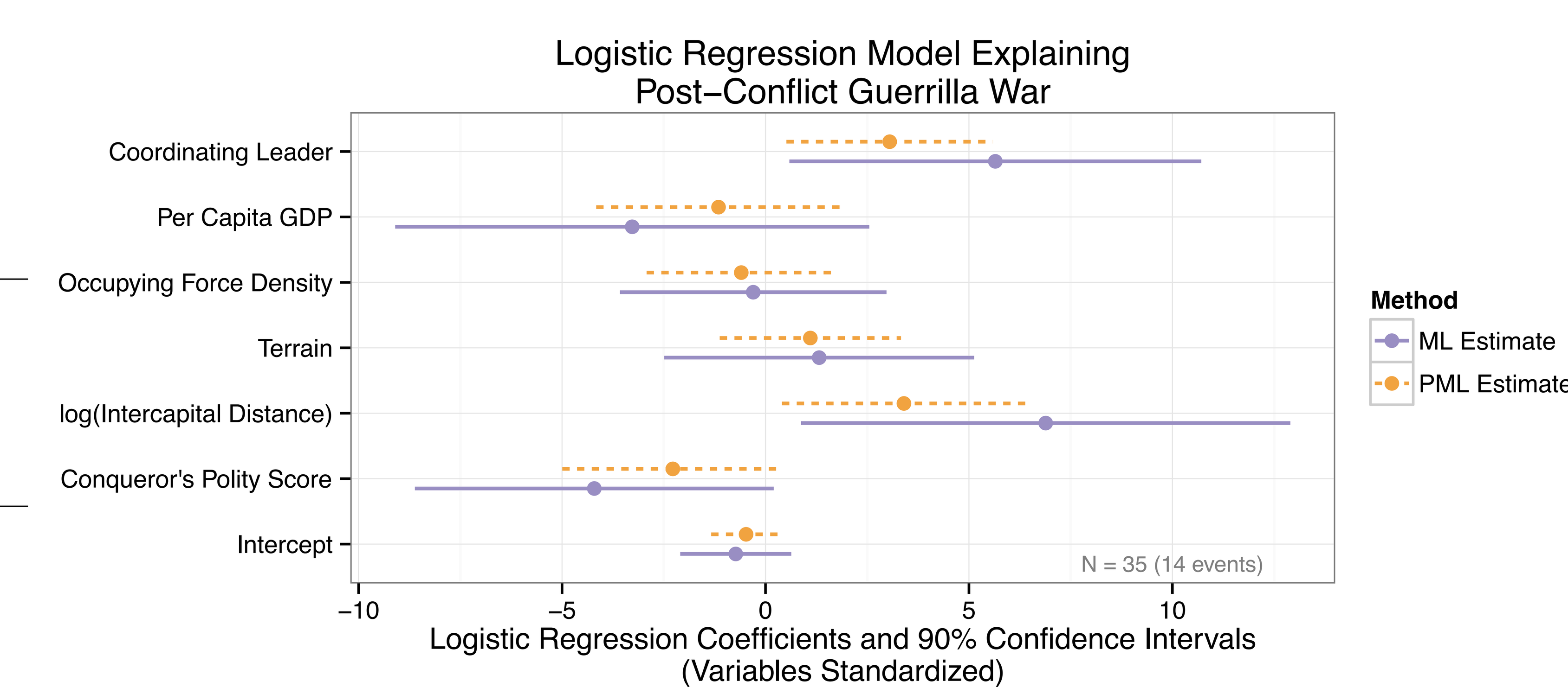
With an N = 35, we use PML to reanalyze this hypothesis and compare these coefficient estimates to those using ML.

Many of the PML estimates are substantially smaller.

- While the coefficient for terrain only changes by 16%, each of the remaining coefficients changes by more than **45%**!

We use out-of-sample prediction to help adjudicate between ML and PML. PML outperforms ML with both Brier and log scoring.

- ML estimates: Brier score of 0.14, log score of 0.58.
- PML estimates: Brier score of 0.12, log score of 0.38.



References

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1): 27-38.

Weisiger, Alex. 2014. "Victory Without Peace: Conquest, Insurgency, and War Termination." *Conflict Management and Peace Science* 31(4): 357-382.