STAT 851 Project - Firth's method for bias reduction of MLEs

Quang Vuong, Nadia Enhaili

26/03/2023

Outline

1. Refresher
2. Motivation
3. Case study
4. Firth's method for bias reduction
5. Example - exponential distribution
6. Application - Firth's method in logistic regression
7. Firth for GLM's in literature
8. Conclusion
9. Questions?

Refresher

$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$

Estimator is biased if it consistently overestimates or underestimates the parameter it is trying to estimate.

$\hat{\theta}_{MLE} = \text{arg max} L(\theta; y_1, ..., y_n)$

Common method for estimating parameters in many statistical models, including linear regression, logistic regression, and many others.

Motivation

- MLEs are consistent and asymptotically unbiased, but there is still a shrinking bias term.
- 1D case:

$$U(\hat{\theta}) = U(\theta) + U'(c(\hat{\theta}))(\hat{\theta} - \theta)$$
$$\Rightarrow \mathbb{E}U'(c(\hat{\theta}))(\hat{\theta} - \theta) = 0.$$

- If U' is not constant, then $\hat{\theta}$ is certainly biased.

Notation

$$\theta = \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^p \end{pmatrix} \in \Omega$$

$$U_r(\theta) = \frac{\partial l(\theta)}{\partial \theta^r} \qquad U_{rs}(\theta) = \frac{\partial^2 l(\theta)}{\partial \theta^r \partial \theta^s}$$

$$\kappa_{r,s} = \frac{1}{n}\mathbb{E} U_r U_s \qquad \kappa_{rs} = \frac{1}{n}\mathbb{E} U_{rs}$$

$$\kappa_{r,s,t} = \frac{1}{n}\mathbb{E} U_r U_s U_t \qquad \kappa_{r,st} = \frac{1}{n}\mathbb{E} U_r U_{st}$$

Motivation

- The first term in the bias of $\hat{\theta}$ has been studied.

$$
\begin{aligned}
\mathbb{E}(\hat{\theta}^r - \theta^r) &= -\frac{\kappa^{r,s}\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})}{2n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \\
&= \frac{1}{n} b_1^r(\theta) + O\left(\frac{1}{n^{\frac{3}{2}}}\right).
\end{aligned}
$$

Case study

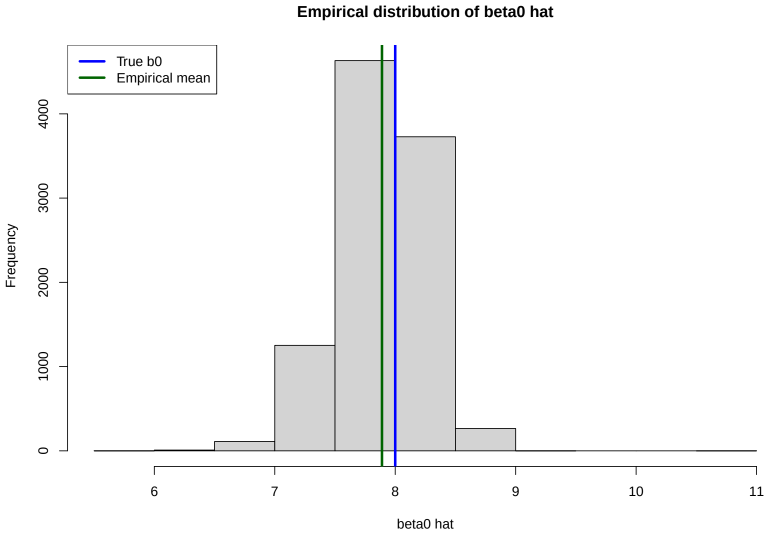
Fit a GLM to data generated by the following model.

$$Y \sim Exp(\lambda)$$

$$\log(\mu) = \log(\frac{1}{\lambda}) = 8 + 5x$$

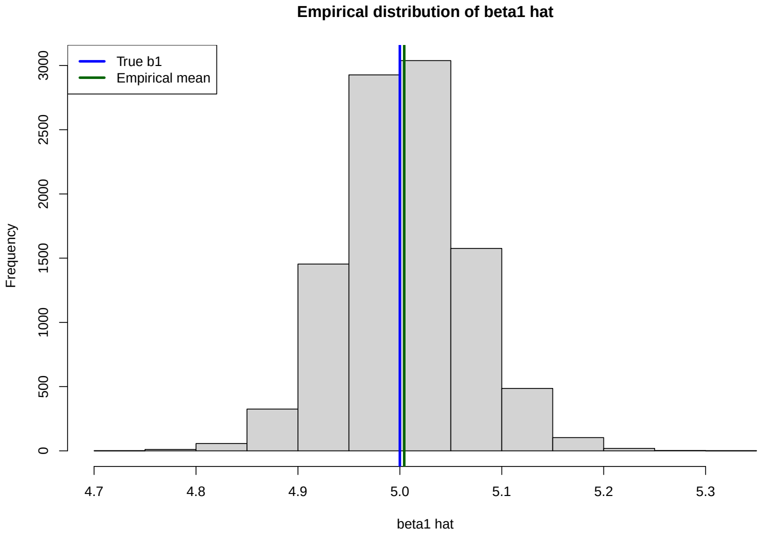
Predictor values x_i are predetermined, while Y_i is repeatedly simulated at x_i 10000 times.

Case study



Empirical distribution of beta0 hat
True b0
Empirical mean
Frequency
0
1000
2000
3000
4000
6
7
8
9
10
11
beta0 hat

Case study


Empirical distribution of beta1 hat
True b1
Empirical mean
Frequency
0
500
1000
1500
2000
2500
3000
4.7
4.8
4.9
5.0
5.1
5.2
5.3
beta1 hat

Firth's method for bias reduction

- Reduce the bias by using an adjusted score function

$$U^*(\theta) = U(\theta) + A(\theta)$$

$$A : \Omega \to \mathbb{R}^p$$

- Root is θ^*

Firth's method for bias reduction

- One can derive the bias of θ^* when estimating θ

$$
\begin{aligned}
\mathbb{E}(\hat{\theta}^r - \theta^r) &= -\frac{\kappa^{r,s}\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})}{2n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \\
&= \frac{1}{n} b_1^r(\theta) + O\left(\frac{1}{n^{\frac{3}{2}}}\right).
\end{aligned}
$$

Firth's method for bias reduction

- Suppose that:

$$
\begin{aligned}
& b_1^r(\theta) + \kappa^{r,s}\mathbb{E}[A^s(\theta)] + O\left(\frac{1}{\sqrt{n}}\right) = 0 \\
\Leftrightarrow\ & \mathbb{E}[A^r(\theta)] = -\kappa_{r,s} b_1^s(\theta) + O\left(\frac{1}{\sqrt{n}}\right) \\
\Leftrightarrow\ & \mathbb{E}[A(\theta)] = -\frac{1}{n}\mathcal{I}(\theta) b_1(\theta) + O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

- Then the first term of the bias is gone.

Firth's method for bias reduction

- Firth (1993) suggests

$$A^{(O)}(\theta) = -\frac{1}{n} I(\theta) b_1(\theta),$$

or

$$A^{(E)}(\theta) = -\frac{1}{n} \mathcal{I}(\theta) b_1(\theta).$$

- The first one leads to more efficient (i.e. less variance) estimators.
- If the model uses an exponential family and θ is the canonical parameter, then $A^{(O)} = A^{(E)}$.

Example - exponential distribution

Let $Y_1, ..., Y_n \sim Exp(\lambda)$. Use Firth's method to reduce the bias of the MLE of λ.

- We have $l(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} Y_i$, so:

$$U(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} Y_i.$$

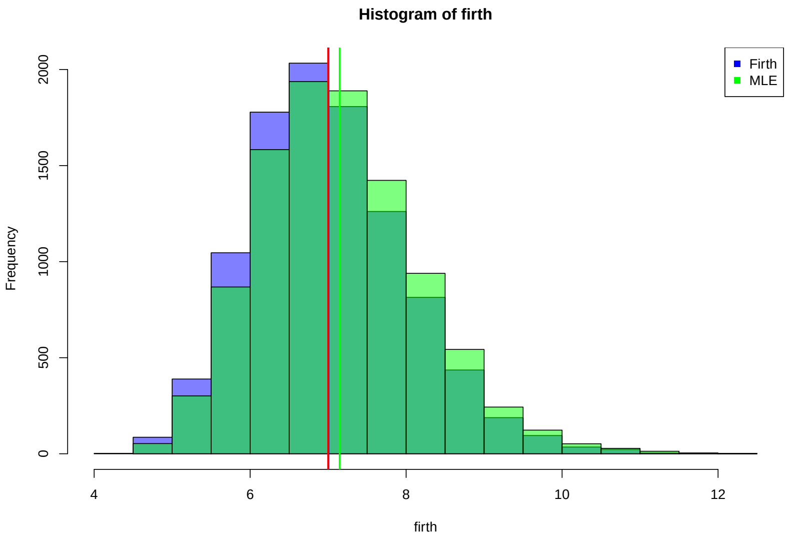
- We can compute $A^{(O)}(\lambda) = A^{(E)}(\lambda) = -\frac{1}{\lambda}$.
- Then

$$U^*(\lambda) = \frac{n-1}{\lambda} - \sum_{i=1}^{n} Y_i$$

and

$$\lambda^* = \frac{n-1}{\sum_{i=1}^{n} Y_i}.$$

Example - exponential distribution



Application - Firth's method in logistic regression

First, let's generate some data based on the following distributions:

$$\beta = (0.5, -0.01, 0.1)$$
$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$X_1 \sim U(1, 150)$$
$$X_2 \sim U(0.5, 100)$$
$$Y \sim Binom(n, \frac{\exp(\eta)}{1+\exp(\eta)})$$

Application (cont'd)

Now we fit a binomial GLM with logit link, as well as a logistic regression model using the R package **logistf**.

```
glm_model <- glm(Y ~ X1 + X2, family = binomial(link
= "logit"), data = sim_data)
```

```
firth_model <- logistf(Y ~ X1 + X2, data = sim_data)
```

We then store the coefficients for each model, and repeat this experiment *n* times.

Application (cont'd) $n = 25$

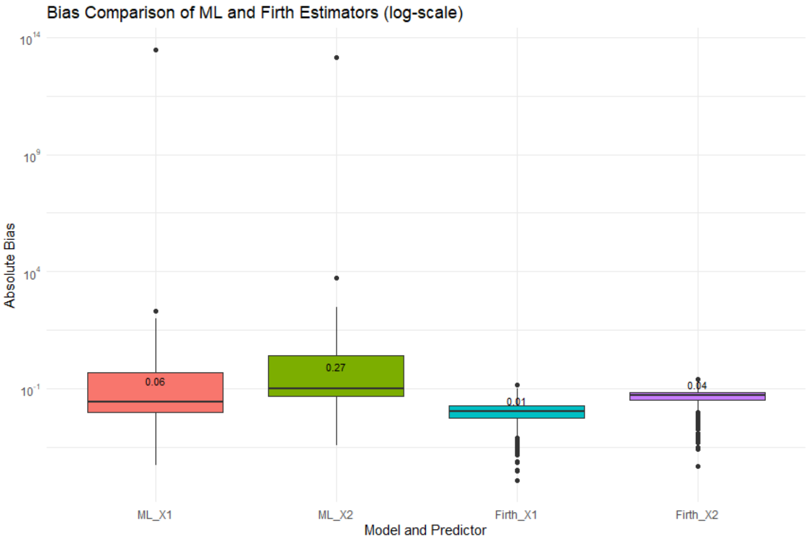


Figure 1: Figure 1

Application (cont'd) $n = 100$

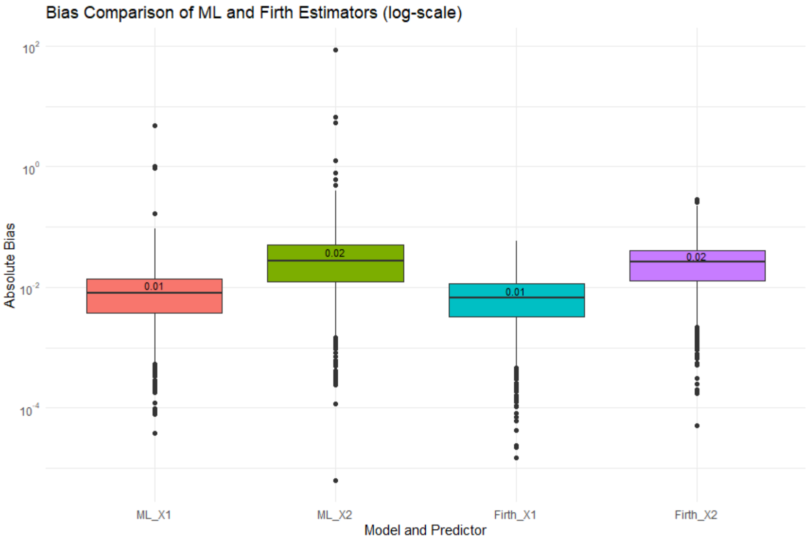


Figure 2: Figure 2

Firth for GLM's in literature

Kosmidis, I., Kenne Pagui, E. C., & Sartori, N. (2020). **Mean and median bias reduction in generalized linear models.**

GLMs with exponential family distributions with dispersion parameter φ and coefficients β.

Section 2.3 proposes an algorithm to solve Firth's modified score equations:

$$\beta^{(j+1)} \leftarrow \beta^{(j)} + \{i_{\beta\beta}^{(j)}\}^{-1} s_{\beta}^{(j)} - b_{\beta}^{(j)}$$
$$\phi^{(j+1)} \leftarrow \phi^{(j)} + \{i_{\phi\phi}^{(j)}\}^{-1} s_{\phi}^{(j)} - b_{\phi}^{(j)}$$

Conclusion

...

Questions?