brglm: Bias reduction in generalized linear models

Ioannis Kosmidis

[1] Department of Statistical Science,
University College London
`ioannis@stats.ucl.ac.uk`

August 2011

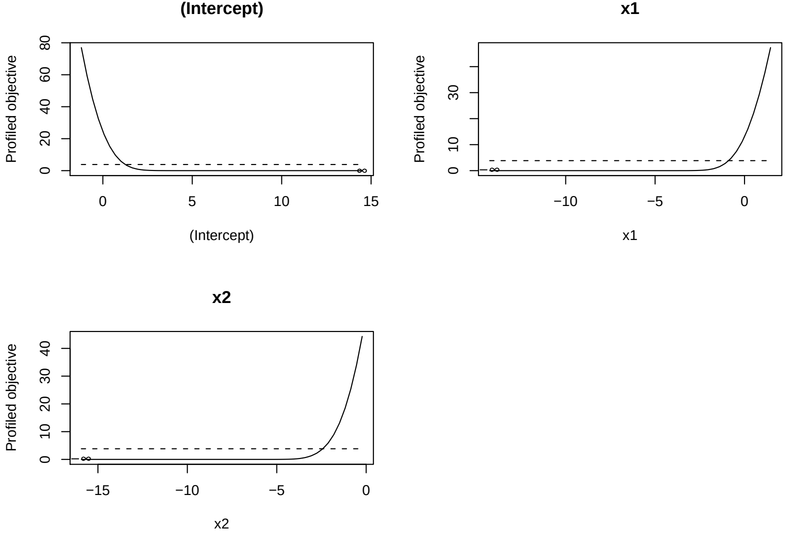
Outline

Infinite maximum likelihood estimates

Data:

X_1	X_2	Successes	Totals
0	0	16	16
	1	1	13
1	0	12	20
	1	0	18

```
> x1 <- c(0, 0, 1, 1)
> x2 <- c(0, 1, 0, 1)
> S <- c(16, 1, 12, 0)
> F <- c(0, 12, 8, 18)
> m1 <- glm(cbind(S, F) ~ x1 + x2,
+         family=binomial(probit))
> coef(m1)
(Intercept)          x1          x2
   6.649437   -6.396090   -8.075514
> coef(summary(m1))[,"Std. Error"]
(Intercept)            x1            x2
   5914.617      5914.617      5914.617
```


(Intercept)
Profiled objective
80
60
40
20
0
0
5
10
15
(Intercept)
x1
Profiled objective
30
10
0
−10
−5
0
x1
x2
Profiled objective
40
30
20
10
0
−15
−10
−5
0
x2

The brglm R package

- Bias-reduced inference in binomial response GLMs via adjusted score function (Firth, 1993; Heinze and Schemper, 2002; K. and Firth, 2009).
- Analysis of sparse settings with binomial responses where the MLE has infinite components or exhibits large bias: e.g.

  → Ingham, S. C. et al. (2010). *Predicting growth-No growth of Listeria monocytogenes on vacuum-packaged ready-to-eat meats.* Journal of Food Protection, 73(4), 708–714.

  → Bell, E. J. L. et al. (2009). *Measuring style with the authorship ratio.* In proceedings of Corpus Linguistics
- In *CRAN* for the past 3 years.
- Pretty stable iteration.
  (assessment biased; contains a bit of personal attachment :)).

Infinite maximum likelihood estimates

```
> ( m2 <- brglm(cbind(S, F) ~ x1 + x2, family =  binomial(probit)) )

Call:  brglm(formula = cbind(S, F) ~ x1 + x2, family = binomial(probit))

Coefficients:
(Intercept)           x1           x2
      1.924       -1.667       -3.165

Degrees of Freedom: 3 Total (i.e. Null);  1 Residual
Deviance:     1.0839  AIC: 12.4378

> confint.default(m2)
                 2.5 %     97.5 %
(Intercept)  0.7272918  3.1216283
x1          -2.9574322 -0.3765631
x2          -4.5887382 -1.7422506

> confint(m2)
Profiling the ordinary deviance for the corresponding ML fit...
Profiling the modified score statistic for the supplied fit...
                2.5 %     97.5 %
(Intercept) 0.8037239        Inf
x1               -Inf -0.4606075
x2               -Inf -1.8017634
```

But...

- `glm` function has changed; since **R** 2.11.0 `glm` handles different fitting methods than `glm.fit` via the `method` argument (`brglm` is just a wrapper function)!
- brglm currently works only with binomial response GLMs.
- New insights on generic unifying algorithms for bias reduction/correction (K. and Firth, 2010).

Models `glm.fit` handles

- Observations on independent random variables $Y_1, \ldots, Y_n$ each with density/mass function of the form

$$f_{Y_i}(y) = \exp\left\{ \frac{y\theta_i - b(\theta_i) - c(y)}{\phi/r_i} - a\left(-\frac{r_i}{\phi}\right) + m(y) \right\} .$$

for some $b(.)$, $c(.)$, $a(.)$ and $m(.)$, with r_i some fixed weight.

$$E(Y_i) = \mu_i = b'(\theta_i)$$

$$Var(Y_i) = \frac{\phi}{w_i} b''(\theta_i) = \frac{\phi}{w_i} V(\mu_i) .$$

- ϕ is a dispersion parameter.
- The mean is associated with a p-dimensional parameter β via

$$g(\mu_i) = \eta_i = \sum_{t=1}^{p} \beta_t x_{it} ,$$

where x_{it} is the $(i,t)th$ component of a model matrix X.

Bias correction

- In regular parametric models the ML estimator $\hat{\gamma}$ is consistent and

$$E\left(\hat{\gamma} - \gamma_0\right) = \frac{b(\gamma_0)}{n} + \frac{b_1(\gamma_0)}{n^2} + \frac{b_2(\gamma_0)}{n^3} + \ldots .$$

Bias-corrective methods

→ Estimate $b(\gamma_0)$ by $b(\hat{\gamma})$.
→ Calculate the bias-corrected estimate $\hat{\gamma} - b(\hat{\gamma})/n$.

Efron (1975) showed that $\hat{\gamma} - b(\hat{\gamma})/n$ has $o(n^{-1})$ bias.

$b(\gamma)/n$ for GLMs

Cordeiro and McCullagh (1991): $b(\gamma)/n = (B_\beta(\beta,\phi), B_\phi(\phi))$

$$B_\beta(\beta,\phi) = \phi(X^TWX)^{-1}X^TW\xi\,,$$

$$B_\phi(\phi) = \frac{(2-p)\phi^3\sum_{i=1}^n r_i a''(-r_i/\phi) - \phi^2\sum_{i=1}^n r_i^3 a'''(-r_i/\phi))}{\left\{\sum_{i=1}^n r_i a''(-r_i/\phi)\right\}^2}\,,$$

where

- $W = \mathrm{diag}\left\{w_1,\ldots,w_n\right\}$, $w_i = r_i d_i^2/V(\mu_i)$ (working weights)
- $\xi = (\xi_1,\ldots,\xi_n)^T$, with $\xi_i = h_i d_i'/(2d_iw_i)$.

Bias reduction

- K. and Firth (2009) show that for GLMs the adjusted score functions for β have the form

$$s_t^*(\beta, \phi) = \frac{1}{\phi} \sum_{i=1}^{n} \frac{w_i}{d_i} \left(y_i + \frac{\phi}{2} h_i \frac{d_i'}{w_i} - \mu_i \right) x_{rt} \quad (t = 1, \ldots, p) \, ,$$

  - $d_i = \mathrm{d}\mu_i/\mathrm{d}\eta_i$, and $d_i' = \mathrm{d}^2\mu_i/\mathrm{d}\eta_i^2$,
  - h_i is the ith "hat" value.
- An expression for the adjusted score function for ϕ is also available.

Bias reduction

- K. and Firth (2009) show that for GLMs the adjusted score functions for β have the form

$$s_t^*(\beta, \phi) = \frac{1}{\phi} \sum_{i=1}^{n} \frac{w_i}{d_i} \left(y_i + \frac{\phi}{2} h_i \frac{d_i'}{w_i} - \mu_i \right) x_{rt} \quad (t = 1, \ldots, p)\,,$$

  - $d_i = \mathrm{d}\mu_i / \mathrm{d}\eta_i$, and $d_i' = \mathrm{d}^2\mu_i / \mathrm{d}\eta_i^2$,
  - h_i is the ith "hat" value.
- An expression for the adjusted score function for ϕ is also available.

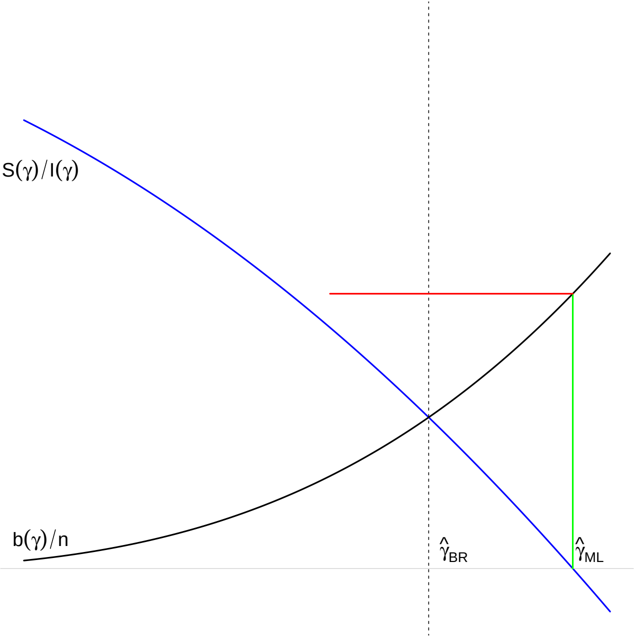
`brglm.fit`'s new iteration

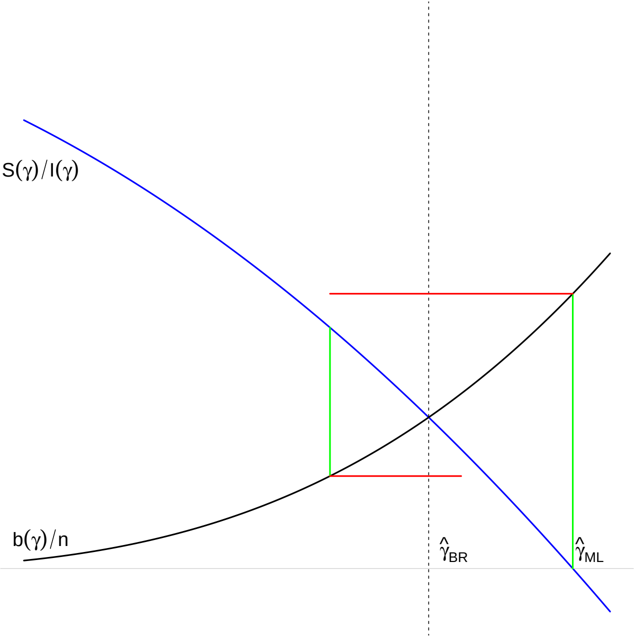
A quasi Fisher-scoring iteration (K. and Firth, 2010)

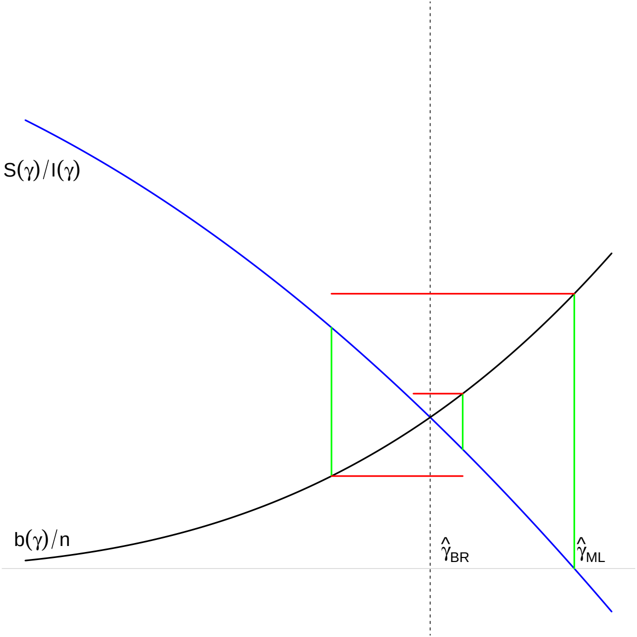
$$\gamma^{(j+1)} := \hat{\gamma}^{(j)} - b(\gamma^{(j)})/n \,,$$

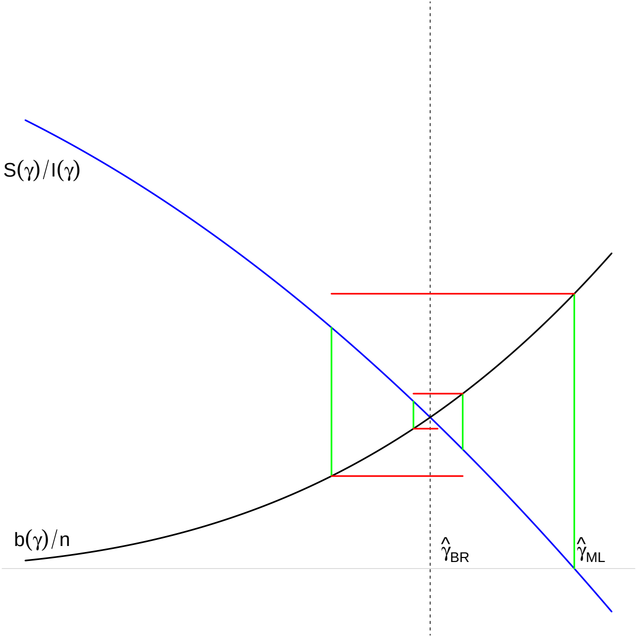
where $\hat{\gamma}^{(j)} = \gamma^{(j)} + \left\{F(\gamma^{(j)})\right\}^{-1} s(\gamma^{(j)})$ (next candidate value for the MLE for β, ϕ from a Fisher-scoring step at $\gamma^{(j)}$.

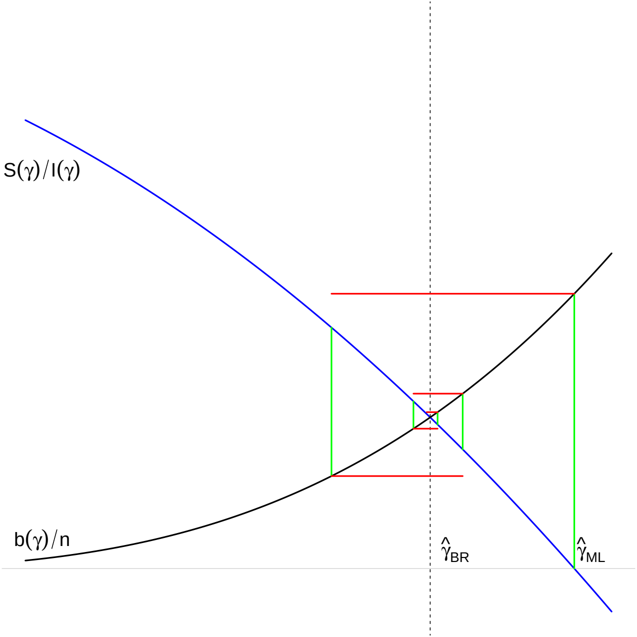
- Bias reduction $\sim$ iterated bias correction

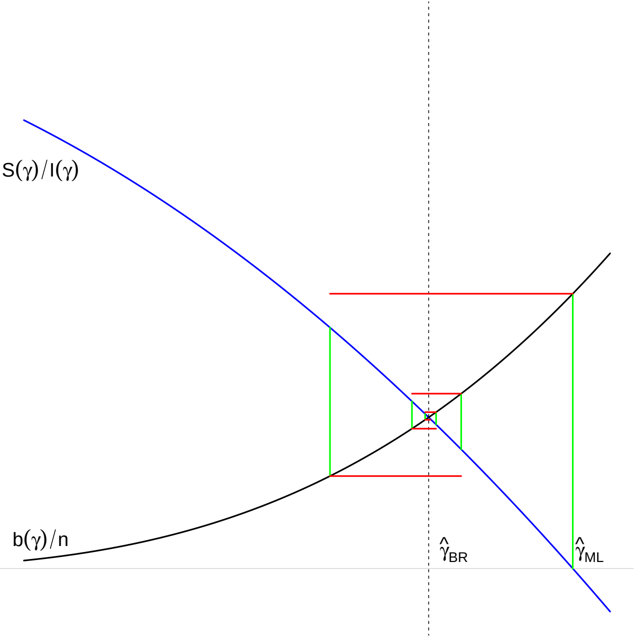

S(γ)/I(γ)
b(γ)/n
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$

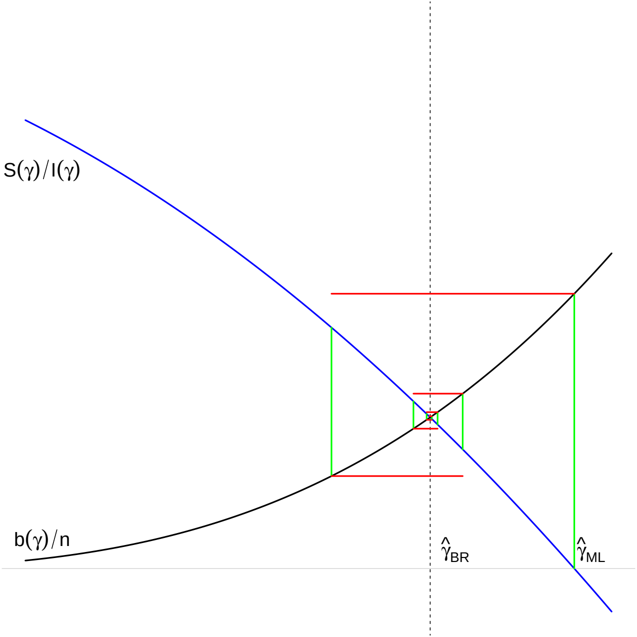

$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$


$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$


$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$


$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$


$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$


$S(\gamma)/I(\gamma)$
$b(\gamma)/n$
$\hat{\gamma}_{BR}$
$\hat{\gamma}_{ML}$

Ingredients

- **R**'s `family` objects provide functions for $d(.)$ (`fam$mu.eta`), for $V(.)$ (`fam$variance`).
- Enrichment of **R**'s `family` objects
  `binomial`, `poisson`, `gaussian`, `Gamma`, `inverse.gaussian`
  to include

Distribution specific	Link specific
$a'(.)$ (`fam$d1afun`)	$d'(.)$ (`fam$dmu.deta`)
$a''(.)$ (`fam$d2afun`)	
$a'''(.)$(`fam$d3afun`)	

brglm.fit: A method for glm

```
> m1

Call:  glm(formula = cbind(S, F) ~ x1 + x2, family = binomial(probit))

Coefficients:
(Intercept)           x1           x2
      6.649       -6.396       -8.076

Degrees of Freedom: 3 Total (i.e. Null);  1 Residual
Null Deviance:      57.7
Residual Deviance: 4.708e-10  AIC: 11.35
> ( m3 <-  update(m1, method = "brglm.fit") )

Call:  glm(formula = cbind(S, F) ~ x1 + x2, family = binomial(probit),
    method = "brglm.fit")

Coefficients:
(Intercept)           x1           x2
      1.924       -1.667       -3.165

Degrees of Freedom: 3 Total (i.e. Null);  1 Residual
Null Deviance:      54.5
Residual Deviance: 1.084  AIC: 12.44
```

Resultant object

```
> class(m3)
[1] "brglm" "glm"     "lm"
```

- Usual methods for `glm` objects apply: `summary`, `coef`, `fitted.values`, `residuals`, `profile` and `confint` (via `profileModel` package), ...

Controlling `brglm.fit`'s iteration

Via the arguments of the `brglm.control` function to be supplied to `glm`'s `control` argument.

- `espilon`: tolerance determining convergence,
- `maxit` : maximum number of iterations allowed,
- `correction`: logical determining whether bias correction or bias reduction is used,
- `dispTrans`: scale at which the bias of ϕ is corrected/reduced: `''identity''`, `''log''`, `''inverse''`, `''sqrt''`, `''custom''`.

A Gamma example

Table: Blood clotting (`example(glm)`).

u	Times
5	118
10	58
15	42
20	35
30	27
40	25
60	21
80	19
100	18

$1/\mu_i = \alpha + \beta \log u_i \ (i = 1, \dots, 9)$

→ uknown shape parameter $1/\phi$.

```
fm1 <- glm(Times ~ log(u),
           family = Gamma,
           method = "brglm.fit",
           correction = FALSE,
           dispTrans = "inverse")
```

Table: Coverage of 95% CIs for β (from 10^5 simulated samples under ML fit)

Type	Est. coverage
Profile Likelihood	91.06
Wald-type (ML)	91.05
Wald-type (BC)	94.76
Wald-type (BR)	94.77

Discussion

- Useful for binomial responses, especially in sparse settings.
- Upcoming version works with all families supported by `glm.fit`

  $$+$$

  multinomial regression (Poisson trick) (K. and Firth, 2011).
- New version will be available soon!

Discussion

- Reducing bias in models with dispersion parameter can result in improved inferential procedures.
  → But not always:

Type	Est. coverage		
	Inverse	Log	Identity
Profile Likelihood	91.06	91.06	91.06
Wald-type (ML)	91.05	91.05	91.05
Wald-type (BC)	94.76	92.07	90.38
Wald-type (BR)	94.77	92.90	91.06

- → Bias relates to parameterization; any attempt to improve it violates exact equivariance under reparameterization. But...

Cordeiro, G. M. and P. McCullagh (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B: Methodological 53*(3), 629–643.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). *The Annals of Statistics 3*, 1189–1217.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine 21*, 2409–2419.

K., I. and D. Firth (2009). Bias reduction in exponential family nonlinear models. *Biometrika 96*(4), 793–804.

K., I. and D. Firth (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics 4*, 1097–1112.

K., I. and D. Firth (2011). Multinomial logit bias reduction via the poisson log-linear model. *Biometrika* (to appear).