Bias reduction in generalized linear models

Ioannis Kosmidis, Eugene Clovis Kenne Pagui and Nicola Sartori

29 April 2017

The brglm2 package

brglm2 provides tools for the estimation and inference from generalized linear models using various methods for bias reduction. Reduction of estimation bias is achieved either through the mean-bias reducing adjusted score equations in Firth (1993) and Kosmidis and Firth (2009), or through the direct subtraction of an estimate of the bias of the maximum likelihood estimator from the maximum likelihood estimates as prescribed in Cordeiro and McCullagh (1991), or through the median-bias reducing adjusted score equations in Kenne Pagui, Salvan, and Sartori (2017).

In the special case of generalized linear models for binomial and multinomial responses, the adjusted score equations approach returns estimates with improved frequentist properties, that are also always finite, even in cases where the maximum likelihood estimates are infinite, like in complete and quasi-complete separation as defined in Albert and Anderson (1984).

The workhorse function is `brglmFit`, which can be passed directly to the `method` argument of the `glm` function. `brglmFit` implements a quasi Fisher scoring procedure, whose special cases result in various explicit and implicit bias reduction methods for generalized linear models (the classification of bias reduction methods into explicit and implicit is given in Kosmidis 2014).

This vignette

This vignette

- presents the supported bias-reducing adjustments to the score functions for generalized linear models
- describes the fitting algorithm at the core of **brglm2**

Other resources

The bias-reducing quasi Fisher scoring iteration is also described in detail in the bias vignette of the **enrichwith** R package. Kosmidis and Firth (2010) describe a parallel quasi Newton-Raphson procedure.

Most of the material in this vignette comes from a presentation by Ioannis Kosmidis at the useR! 2016 international R User conference at University of Stanford on 16 June 2016. The presentation was titled "Reduced-bias inference in generalized linear models" and can be watched online at this link.

Generalized linear models

Model

Suppose that $y_1, \ldots, y_n$ are observations on independent random variables $Y_1, \ldots, Y_n$, each with probability density/mass function of the form

$$f_{Y_i}(y) = \exp\left\{\frac{y\theta_i - b(\theta_i) - c_1(y)}{\phi/m_i} - \frac{1}{2}a\left(-\frac{m_i}{\phi}\right) + c_2(y)\right\}$$

for some sufficiently smooth functions $b(.)$, $c_1(.)$, $a(.)$ and $c_2(.)$, and fixed observation weights $m_1, \ldots, m_n$. The expected value and the variance of Y_i are then

$$
\begin{aligned}
E(Y_i) &= \mu_i = b'(\theta_i) \\
Var(Y_i) &= \frac{\phi}{m_i} b''(\theta_i) = \frac{\phi}{m_i} V(\mu_i)
\end{aligned}
$$

Hence, in this parameterization, ϕ is a dispersion parameter.

A generalized linear model links the mean μ_i to a linear predictor η_i as

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

where $g(.)$ is a monotone, sufficiently smooth link function, taking values on $\Re$, x_{it} is the $(i,t)th$ component of a model matrix X, and $\beta = (\beta_1, \ldots, \beta_p)^\top$.

Score functions and information matrix

Suppressing the dependence of the various quantities on the model parameters and the data, the derivatives of the log-likelihood about β and ϕ (score functions) are

$$
\begin{aligned}
s_\beta &= \frac{1}{\phi} X^T W D^{-1} (y - \mu) \\
s_\phi &= \frac{1}{2\phi^2} \sum_{i=1}^{n} (q_i - \rho_i)
\end{aligned}
$$

with $y = (y_1, \ldots, y_n)^\top$, $\mu = (\mu_1, \ldots, \mu_n)^\top$, $W = \operatorname{diag}\{w_1, \ldots, w_n\}$ and $D = \operatorname{diag}\{d_1, \ldots, d_n\}$, where $w_i = m_i d_i^2 / v_i$ is the ith working weight, with $d_i = d\mu_i / d\eta_i$ and $v_i = V(\mu_i)$. Furthermore, $q_i = -2m_i\{y_i\theta_i - b(\theta_i) - c_1(y_i)\}$ and $\rho_i = m_i a'_1$ are the ith deviance residual (e.g. as is implemented in the `dev.resids` component of most `family` objects) and its expectation, respectively, with $a'_i = a'(-m_i/\phi)$. The only `family` object deviating from the above description is `Gamma` where `Gamma()$dev.resids` implements $q_i - 2m_i$ instead of q_i. For convenience in implementation, and just for Gamma we define $\rho_i = m_i a'_1 - 2m_i = -2\psi(-m_i/\phi) + 2\log(-m_i/\phi)$, where ψ is the `digamma` function. This change affects none of the estimation methods discussed in this vignette.

The expected information matrix about β and ϕ is

$$
i = \begin{bmatrix} i_{\beta\beta} & 0_p \\ 0_p^\top & i_{\phi\phi} \end{bmatrix} = \begin{bmatrix} \frac{1}{\phi} X^\top W X & 0_p \\ 0_p^\top & \frac{1}{2\phi^4} \sum_{i=1}^{n} m_i^2 a''_i \end{bmatrix},
$$

where 0_p is a p-vector of zeros, and $a''_i = a''(-m_i/\phi)$.

Maximum likelihood estimation

The maximum likelihood estimators $\hat{\beta}$ and $\hat{\phi}$ of β and ϕ, respectively, can be found by the solution of the score equations $s_\beta = 0_p$ and $s_\phi = 0$.

Mean bias-reducing adjusted score functions

Let $A_\beta = -i_\beta b_\beta$ and $A_\phi = -i_\phi b_\phi$, where b_β and b_ϕ are the first terms in the expansion of the mean bias of the maximum likelihood estimator of the regression parameters β and dispersion ϕ, respectively. The results in Firth (1993) can be used to show that the solution of the adjusted score equations

$$
\begin{aligned}
s_\beta + A_\beta &= 0_p \\
s_\phi + A_\phi &= 0
\end{aligned}
$$

results in estimators $\tilde\beta$ and $\tilde\phi$ with bias of smaller asymptotic order than the maximum likelihood estimator.

The results in either Kosmidis and Firth (2009) or Cordeiro and McCullagh (1991) can then be used to re-express the adjustments in forms that are convenient for implementation. In particular, and after some algebra the bias-reducing adjustments for generalized linear models are

$$
\begin{aligned}
A_\beta &= X^\top W \xi\,, \\
A_\phi &= \frac{(p-2)}{2\phi} + \frac{\sum_{i=1}^n m_i^3 a_i'''}{2\phi^2 \sum_{i=1}^n m_i^2 a_i''}
\end{aligned}
$$

where $\xi = (\xi_1, \ldots, \xi_n)^T$ with $\xi_i = h_i d_i'/(2 d_i w_i)$, $d_i' = d^2\mu_i/d\eta_i^2$, $a_i'' = a''(-m_i/\phi)$, $a_i''' = a'''(-m_i/\phi)$, and h_i is the "hat" value for the ith observation (see, e.g. `?hatvalues`).

Median bias-reducing adjusted score functions

The results in Kenne Pagui, Salvan, and Sartori (2017) can be used to show that if

$$
\begin{aligned}
A_\beta &= X^\top W(\xi + Xu) \\
A_\phi &= \frac{p}{2\phi} + \frac{\sum_{i=1}^n m_i^3 a_i'''}{6\phi^2 \sum_{i=1}^n m_i^2 a_i''}\,,
\end{aligned}
$$

then the solution of the adjusted score equations $s_\beta + A_\beta = 0_p$ and $s_\phi + A_\phi = 0$ results in estimators $\tilde\beta$ and $\tilde\phi$ with median bias of smaller asymptotic order than the maximum likelihood estimator. In the above expression, $u = (u_1, \ldots, u_p)^\top$ with

$$
u_j = [(X^\top W X)^{-1}]_j^\top X^\top \begin{bmatrix} \tilde h_{j,1}\left\{d_1 v_1'/(6v_1) - d_1'/(2d_1)\right\} \\ \vdots \\ \tilde h_{j,n}\left\{d_n v_n'/(6v_n) - d_n'/(2d_n)\right\} \end{bmatrix}
$$

where $[A]_j$ denotes the jth row of matrix A as a column vector, $v_i' = V'(\mu_i)$, and $\tilde h_{j,i}$ is the ith diagonal element of XK_jX^TW, with $K_j = [(X^\top W X)^{-1}]_j [(X^\top W X)^{-1}]_j^\top / [(X^\top W X)^{-1}]_{jj}$.

Fitting algorithm in `brglmFit`

`brglmFit` implements a quasi Fisher scoring procedure for solving the adjusted score equations $s_\beta + A_\beta = 0_p$ and $s_\phi + A_\phi = 0$. The iteration consists of an outer loop and an inner loop that implements step-halving. The algorithm is as follows:

Input

- s_β, $i_{\beta\beta}$, A_β
- s_ϕ, $i_{\phi\phi}$, A_ϕ
- Starting values $\beta^{(0)}$ and $\phi^{(0)}$
- $\epsilon > 0$: tolerance for the $L1$ norm of the direction before reporting convergence
- M: maximum number of halving steps that can be taken

Output

- $\tilde{\beta}$, $\tilde{\phi}$

Iteration

Initialize outer loop

1. $k \leftarrow 0$
2. $\upsilon_\beta^{(0)} \leftarrow \left\{i_{\beta\beta}\left(\beta^{(0)}, \phi^{(0)}\right)\right\}^{-1}\left\{s_\beta\left(\beta^{(0)}, \phi^{(0)}\right) + A_\beta\left(\beta^{(0)}, \phi^{(0)}\right)\right\}$
3. $\upsilon_\phi^{(0)} \leftarrow \left\{i_{\phi\phi}\left(\beta^{(0)}, \phi^{(0)}\right)\right\}^{-1}\left\{s_\phi\left(\beta^{(0)}, \phi^{(0)}\right) + A_\phi\left(\beta^{(0)}, \phi^{(0)}\right)\right\}$

Initialize inner loop

4. $m \leftarrow 0$
5. $b^{(m)} \leftarrow \beta^{(k)}$
6. $f^{(m)} \leftarrow \phi^{(k)}$
7. $\upsilon_\beta^{(m)} \leftarrow \upsilon_\beta^{(k)}$
8. $\upsilon_\phi^{(m)} \leftarrow \upsilon_\phi^{(k)}$
9. $d \leftarrow \left|\upsilon_\beta^{(m)}\right|_1 + \left|\upsilon_\phi^{(m)}\right|$

Update parameters

10. $b^{(m+1)} \leftarrow b^{(m)} + 2^{-m}\upsilon_\beta^{(m)}$
11. $f^{(m+1)} \leftarrow f^{(m)} + 2^{-m}\upsilon_\phi^{(m)}$

Update direction

12. $\upsilon_\beta^{(m+1)} \leftarrow \left\{i_{\beta\beta}\left(b^{(m+1)}, f^{(m+1)}\right)\right\}^{-1}\left\{s_\beta\left(b^{(m+1)}, f^{(m+1)}\right) + A_\beta\left(b^{(m+1)}, f^{(m+1)}\right)\right\}$
13. $\upsilon_\phi^{(m+1)} \leftarrow \left\{i_{\phi\phi}\left(b^{(m+1)}, f^{(m+1)}\right)\right\}^{-1}\left\{s_\phi\left(b^{(m+1)}, f^{(m+1)}\right) + A_\phi\left(b^{(m+1)}, f^{(m+1)}\right)\right\}$

Continue or break halving within inner loop

14. if $m + 1 < M$ and $\left|\upsilon_\beta^{(m+1)}\right|_1 + \left|\upsilon_\phi^{(m+1)}\right| > d$
    14.1. $m \leftarrow m + 1$
    14.2. GO TO 10
15. else
    15.1. $\beta^{(k+1)} \leftarrow b^{(m+1)}$
    15.2. $\phi^{(k+1)} \leftarrow f^{(m+1)}$

15.3. $v_\beta^{(k+1)} \leftarrow v_b^{(m+1)}$

15.4. $v_\phi^{(k+1)} \leftarrow v_f^{(m+1)}$

Continue or break outer loop

16. if $k+1 < K$ and $\left|v_\beta^{(k+1)}\right|_1 + \left|v_\phi^{(k+1)}\right| > \epsilon$

    16.1 $k \leftarrow k+1$

    16.2. GO TO 4

17. else

    17.1. $\tilde\beta \leftarrow \beta^{(k+1)}$

    17.2. $\tilde\phi \leftarrow \phi^{(k+1)}$

    17.3. STOP

Notes

- For $K = M = 1$, $\beta^{(0)} = \hat\beta$ and $\phi^{(0)} = \hat\phi$, the above iteration computes the bias-corrected estimates proposed in Cordeiro and McCullagh (1991). This is achieved with the `brglmFit` method is used with `type = "correction"` (see `?brglmFit`).
- The mean-bias reducing adjusted score functions are solved when the `brglmFit` method is used with `type = "AS_mean"`, and the median-bias reducing adjusted score functions with `type = AS_median` (see `?brglmFit`).
- The steps where ϕ and the ϕ direction are updated are ignored for generalized linear models with known dispersion parameter, like in models with binomial and poisson responses. Also, in that case, $v_\phi^{(.)}$ and $v_\phi^{(.)}$ in steps 9, 14 and 16 are set to zero.
- The implementation of the adjusted score functions requires ready implementations of $d^2\mu_i/d\eta_i^2$, $a'(.)$, $a''(.)$ and $a'''(.)$. The **enrichwith** R package is used internally to enrich the base `family` and `link-glm` objects with implementations of those functions (see `?enrich.family` and `?enrich.link-glm`).
- The above iteration can be used to implement a variety of additive adjustments to the score function, by supplying the algorithm with appropriate adjustment functions A_β and A_ϕ

Contributions to this vignette

The first version of the vignette has been written by Ioannis Kosmidis. Eugene Clovis Kenne Pagui and Nicola Sartori contributed the first version of the section "Median bias-reducing adjusted score functions", and Ioannis Kosmidis brought the expressions for the median bias-reducing adjustments in the reduced form that is shown above and is implemented in `brglmFit`.

References

Albert, A., and J. A. Anderson. 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 71 (1): 1–10.

Cordeiro, G. M., and P. McCullagh. 1991. "Bias Correction in Generalized Linear Models." *Journal of the*

Royal Statistical Society, Series B: Methodological 53 (3): 629–43.

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

Kenne Pagui, E. C., A. Salvan, and N. Sartori. 2017. "Median Bias Reduction of Maximum Likelihood Estimates." *ArXiv E-Prints*. http://arxiv.org/abs/1604.04768.

Kosmidis, I. 2014. "Bias in Parametric Estimation: Reduction and Useful Side-Effects." *Wiley Interdisciplinary Reviews: Computational Statistics* 6 (3). John Wiley & Sons, Inc.: 185–96. doi:10.1002/wics.1296.

Kosmidis, I., and D. Firth. 2009. "Bias Reduction in Exponential Family Nonlinear Models." *Biometrika* 96 (4): 793–804. doi:10.1093/biomet/asp055.

———. 2010. "A Generic Algorithm for Reducing Bias in Parametric Estimation." *Electronic Journal of Statistics* 4: 1097–1112. doi:10.1214/10-EJS579.