

Calibrated h-likelihood with bias correction in SAE (hglm-bc) vignette

Nirosha Rathnayake*, Dai (Daisy) Hongying†

Contents

0.0.1	Introduction	1
0.0.2	Proposed h-likelihood with bias correction in small area estimation	2

`library(hglm-bc2)`

0.0.1 Introduction

The `hglm-bc` package vignette illustrates a series of applications that may be of interest in Small Area Estimation (SAE) based on the calibrated h -likelihood approach. In SAE a random effect or the latent variable u_i describes the between area estimation. The proposed calibrated hierarchical h -likelihood in SAE obtains SAE through hierarchical estimation of **fixed effects** and **random effects** with bias correction using Regression Calibration Method (RCM). This approach allows the conditional distribution of y given random effect (u) from any exponential family distributions (GLM family) and the random effect $u \sim N(0, \sigma^2)$ [3]. This approach is called the *calibrated h-likelihood with bias correction*. This `hglm-bc` is mainly built to fit exponential family distributions (GLM family) via h -likelihood with *Bias - Correction* in SAE. This package also can be applied to *Hierarchical Generalized Linear Models (HGLM)* where $y|u$ from the exponential family distribution and u may also be from the exponential family distribution, *without bias correction approach* [1, 2]. The CPH with bias correction approach is described in detail in the PhD dissertation report, (See the article). The h -likelihood is the sum of conditional log likelihood of $y|u$ and the log likelihood of u ,

$$h = \sum_{i=1}^m \sum_{j=1}^{n_i} \ell_{y_{ij}|u_i} + \sum_{i=1}^m \ell_{u_i},$$

- The h -likelihood for the Binomial-Normal HGLM (Mixed Logit Model) can be written as

$$h = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij}(x_{ij}^T \beta + u_i) - \log(1 + \exp(x_{ij}^T \beta + u_i))) - \sum_{i=1}^m \frac{m}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2 - \frac{m}{2} \log(2\pi)$$

The estimated model parameters through the proposed method are called the maximum hierarchical likelihood estimators (MHLEs). Let $\tau = (\beta, u)$ be the unknown parameters. Given the variance parameter $\theta = (\theta_1, \dots, \theta_q)^T = \sigma^2$, MHLE of $\hat{\tau} = (\hat{\beta}, \hat{u})$ are obtained by solving the score function $\partial h / \partial \tau = 0$. When

*University of Nebraska Medical Center, niro.uno@gmail.com, <https://niroshar.github.io/My-Profile/>

†University of Nebraska Medical Center, <https://www.unmc.edu/publichealth/departments/biostatistics/facultyandstaff/hongying-dai.html>

the solution does not have a closed form, the MHLE for (β, u) are obtained using *Newton-Raphson* approximation through an iterative procedure using equation. Unlike in the standard linear mixed model, the model (1) does not have a closed-form for the joint log likelihood, hence it is very challenging to estimate BLUP or EBLUP of model parameters. In such situations, h -likelihood plays a major role in simplifying the parameter estimation procedure. Mostly, the MHLEs are often obtained via numerical approximation methods due to intractable integrals in the joint log-likelihood function.

The score function \mathcal{S} and the hessian matrix \mathcal{J} of the Newton Raphson iterative algorithm shown below,

$$\mathcal{S} = \begin{pmatrix} \frac{\partial h}{\partial \beta_r} \\ \frac{\partial h}{\partial u_i} \end{pmatrix}, \mathcal{J} = \begin{pmatrix} -\frac{\partial^2 h}{\partial \beta_r \partial \beta_s} & -\frac{\partial^2 h}{\partial \beta_r \partial u_i} \\ -\frac{\partial^2 h}{\partial u_i \partial \beta_s} & -\frac{\partial^2 h}{\partial u_i \partial u_l} \end{pmatrix}$$

The estimated \mathcal{J} is the asymptotic variance-covariance matrix of $\hat{\beta}$ and \hat{u} . The variance-covariance matrix for variance component $\hat{\theta}$ will be estimated by MHLE of θ using h_A through an iterative procedure.

0.0.1.1 Bias Correction of Random Effects we consider bias correction of random effects within the parameter estimation process to mitigate the biasness that could occur due to the use of current estimates of random effects ($\hat{u}^{(t)}$) to obtain estimations for variance parameter $\theta(\sigma^2)$. The MHLE of (\hat{u}) will lead to bias and inconsistent estimations of regression coefficients and parameters in the random effects, such as variance parameters. We apply the *Regression Calibration Method (RCM)* which corrects the biasness in the estimators introducing a correction factor to adjust the current estimators based on the variance of the previous and current estimators of u . Under RCM, the current estimate of u, \hat{u} is replaced by the expectation of conditional expectation of $u|\hat{u}$. Given $u \sim N(0, \sigma^2)$,

$$E[u|\hat{u}] = \zeta \hat{u} = \frac{\sigma^2}{\sigma^2 + \gamma^2} \hat{u},$$

$$E[\exp(u)|\hat{u}] = \exp\left(\zeta \hat{u} + \frac{\sigma^2(1 - \zeta)}{2}\right),$$

where $\zeta = \sigma^2/(\sigma^2 + \gamma^2)$ is the correction factor, and $\gamma = \text{diag}(\Sigma_{\hat{u}})$, $\Sigma_{\hat{u}} = (\mathcal{J}^{-1})_{22}$ is a vector of m elements obtained from the variance-covariance matrix of \hat{u} . The dispersion parameters are obtained after correcting to the current \hat{u} . The dispersion parameters will be estimated iteratively based on the adjusted h -likelihood function.

0.0.2 Proposed h-likelihood with bias correction in small area estimation

The proposed ‘calibrated h-likelihood with bias correction’ approach is currently limited to GLM family distributions where a response variable whose conditional distribution of $y|u$ belongs to exponential family and the random effect $u \sim N(0, \sigma^2)$. The algorithm consists of two main key components:

1. A linear predictor: $\eta_i = \sum_{i=1}^n x_i \beta_i$,
2. A link function: $\eta_i = g(\mu_i)$,

The list of components in each GLM family distribution in R is illustrated in an object of class, `family(run ?family` in R console for more details). As an example, the code below shows the constituent parts for the binomial GLM, which is what is used to fit linear logistic regression:

```

b.family <- binomial()
class(b.family)
#> [1] "family"
names(b.family)
#> [1] "family"      "link"      "linkfun"    "linkinv"    "variance"
#> [6] "dev.resids"  "aic"      "mu.eta"     "initialize" "validmu"
#> [11] "valideta"    "simulate"

p.family <- poisson()

p.family
#>
#> Family: poisson
#> Link function: log
b.family
#>
#> Family: binomial
#> Link function: logit

```

Example: Binomial-Normal HGLM (Mixed Logit Model) with Bias Correction Binomial-Normal HGLM is also known as the mixed logit model in GLM family with the binary response variable and the random effect $u \sim N(0, \sigma^2)$.

```

library(hglm)
#>
#> Attaching package: 'hglm'
#> The following objects are masked from 'package:hglm2':
#>
#>     everysmoke, hglm
## basic example code
data <- everysmoke
mformula <- "smoke_ever ~ as.factor(age) + as.factor(gender) + as.factor(race) + as.factor(year) + povt.
dom <- "county"
y.family <- "binomial"
rand.family <- "gaussian"
## Fit the model
# hglm.fit <- hglm(data=everysmoke, mformula, dom = "county", y.family=binomial)
# hglm.fit$summary

```

References

- [1] Youngjo Lee and John A Nelder. “Hierarchical generalized linear models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.4 (1996), pp. 619–656.
- [2] Youngjo Lee, John A Nelder, and Yudi Pawitan. *Generalized linear models with random effects: unified analysis via H-likelihood*. Chapman and Hall/CRC, 2006.
- [3] Charles E McCulloch and John M Neuhaus. “Generalized linear mixed models”. In: *Encyclopedia of biostatistics* 4 (2005).