

2021-09-22

# A PolyaGamma sampler for a generalized logistic regression

Dalla Valle, Luciana

<http://hdl.handle.net/10026.1/15149>

---

10.1080/00949655.2021.1910947

Journal of Statistical Computation and Simulation

Informa UK Limited

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

# A Pólya-Gamma Sampler for a Generalized Logistic Regression

Luciana Dalla Valle\*

University of Plymouth, Plymouth, UK

Fabrizio Leisen

University of Kent, Canterbury, UK

Luca Rossini

VU University Amsterdam, Amsterdam, Netherlands

and

Weixuan Zhu

Xiamen University, Xiamen, People's Republic of China

September 9, 2019

## Abstract

In this paper we introduce a novel Bayesian data augmentation approach for estimating the parameters of the generalised logistic regression model. We propose a Pólya-Gamma sampler algorithm that allows us to sample from the exact posterior distribution, rather than relying on approximations. A simulation study illustrates the flexibility and accuracy of the proposed approach to capture heavy and light tails in binary response data of different dimensions. The methodology is applied to two different real datasets, where we demonstrate that the Pólya-Gamma sampler provides more precise estimates than the empirical likelihood method, outperforming approximate approaches.

*Keywords:* Bayesian inference; generalized logistic regression; Pólya-Gamma sampler; recidivism data.

---

\*The authors gratefully acknowledge Alex Diana for advising on the PGdraw sampler. The second author was supported by the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no: 630677. The third author acknowledges financial support from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 796902. The fourth author is supported by the Chinese Fundamental Research Funds for the Central Universities No 20720181062.

# 1 Introduction

Data augmentation (DA) is undoubtedly one of the most popular Markov Chain Monte Carlo (MCMC) methods. The idea is to consider the target distribution as the marginal of a joint distribution in an augmented space. Assuming it is easy to sample from the conditional distributions, the algorithm is a straightforward two steps procedure which makes use of a simple Gibbs sampling. Since the seminal work of Tanner and Wong (1987), data augmentation has been extensively studied, developed and used in different contexts. For instance, Meng and van Dyk (1999), Liu and Wu (1999), van Dyk and Meng (2001) developed strategies to speed up the basic DA algorithm. Hobert and Marchev (2008) studied the efficiency of DA algorithms. Recently, Leisen et al. (2017) use a DA approach to perform Bayesian inference for the Yule-Simon distribution. The idea is used in the context of infinite hidden Markov models by Hensley and Djurić (2017). The literature on DA is vast and it is difficult to list all the contributions to the field. A review of DA algorithms can be found in Brooks et al. (2011), Chapter 10.

In the context of binary regression, the article of Albert and Chib (1993) sets a milestone in the use of DA algorithms. The authors develop exact Bayesian methods for modeling categorical response data by introducing suitable auxiliary variables. In particular, they use a DA approach to estimate the parameters of the probit regression model. However, although some valid approaches have been proposed (Holmes et al., 2006), Bayesian inference for the logistic regression was not satisfactorily addressed until the paper of Polson et al. (2013). The authors proposed an elegant DA approach which makes use of the following identity:

$$\frac{(e^\psi)^a}{(1 + e^\psi)^b} = 2^{-b} e^{\kappa\psi} \int e^{-\omega\psi^2/2} p(\omega) d\omega \quad (1)$$

where  $\kappa = a - b/2$ ,  $a > 0$ ,  $b > 0$  and  $p(\omega)$  is the density of a  $\text{PG}(b,0)$  distribution. The distribution  $\text{PG}(b,0)$  is a Pólya-Gamma distribution with parameters  $b$  and  $0$ ; we refer to the

article of Polson et al. (2013) for details about Pólya-Gamma distributions. When  $\psi = \mathbf{x}^T \boldsymbol{\beta}$  is a linear function of predictors, the integrand is the kernel of a Gaussian likelihood in  $\boldsymbol{\beta}$ . This naturally suggests a DA scheme which simply requires to be able to sample from the multivariate normal and the Pólya-Gamma distributions. This algorithm is called the *Pólya-Gamma sampler*. A theoretical study of the algorithm can be found in Choi and Hobert (2013).

In this paper, we propose a Pólya-Gamma sampler to estimate the parameters of the generalized logistic regression model introduced in Dalla Valle et al. (2019). The authors studied immigration in Europe employing a novel regression model which makes use of a generalized logistic distribution. Their work is motivated by the need of a logistic model with heavy tails. The distribution considered in the paper has the following density function:

$$f(x) = \frac{1}{B(p, p)} \frac{e^{px}}{(1 + e^x)^{2p}} \quad x \in \mathbb{R}. \quad (2)$$

The parameter  $p$  controls the tails of the distribution. The standard logistic distribution can be recovered with  $p = 1$ . Compared with the standard logistic distribution,  $p > 1$  means lighter tails and  $0 < p < 1$  means heavier tails, as shown in Figure 1. Unfortunately, the distribution function is not explicit:

$$F(x) = \frac{1}{B(p, p)} B\left(\frac{e^x}{1 + e^x}; p, p\right) \quad x \in \mathbb{R} \quad (3)$$

where  $B(t; p, p) = \int_0^t x^{p-1}(1-x)^{p-1} dx$ , with  $0 < t < 1$ , is the incomplete Beta function. Therefore, a straightforward use of the methodology of Polson et al. (2013) is not possible. To overcome the problem, Dalla Valle et al. (2019) use an approximate Bayesian computational method that relies on the empirical likelihood (see Mengersen et al. (2013) and Karabatsos and Leisen (2018)). However, since the empirical likelihood approach is based on an approximation of the posterior distribution, this method may lead to wide credible intervals and sometimes ambiguous estimates, especially for the tail parameter.

In this paper, we propose a novel approach for estimating the generalised logistic regression, that allows us to draw samples from the exact posterior distribution and is able to overcome the drawbacks of the empirical likelihood method. We show how to set a DA scheme for the generalized logistic regression which makes use of the Pólya-Gamma identity in equation (1). We test the performance of the proposed approach on simulated data and on two different real dataset: the first dataset regards people’s opinions towards immigration in Europe; the second contains information about the recidivism of criminals detained in Iowa. We show that the Pólya-Gamma sampler yields accurate results in high-dimension and outperforms the empirical likelihood algorithm providing a more precise estimation of the model parameters.

The rest of the paper is organised as follows. In Section 2, we introduce the generalized logistic regression framework in a Bayesian setting. In Section 3, we provide the details of the Pólya-Gamma sampler for this model. Section 4 and Section 5 illustrate the algorithm performance with simulated and real data. Section 6 concludes.

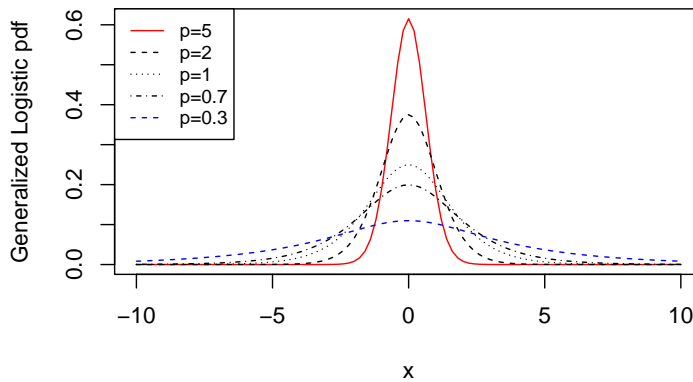


Figure 1: Probability density function of the generalized logistic with  $p = 1$ ,  $p = 0, 3$ ,  $p = 0, 7$ ,  $p = 2$  and  $p = 5$ .

## 2 The Generalized Bayesian Logistic Regression

Consider a binary regression set-up in which  $Y_1, \dots, Y_n$  are independent Bernoulli random variables such that  $\Pr(Y_i = 1 | \boldsymbol{\beta}) = H(x_i^T \boldsymbol{\beta})$  where  $x_i$  is a  $k \times 1$  vector of known covariates associated with  $Y_i$ ,  $i = 1, \dots, n$ ,  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of unknown regression coefficients, and  $H : \mathbb{R} \rightarrow (0, 1)$  is a distribution function. In this case, the likelihood function is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [H(x_i^T \boldsymbol{\beta})]^{y_i} [1 - H(x_i^T \boldsymbol{\beta})]^{1-y_i} \quad (4)$$

If  $H(x) = \Phi(x)$  is the distribution function of a Gaussian, then we are in the probit regression framework. If  $H(x) = e^x(1 + e^x)^{-1}$ , then we are in the special case of the standard logistic regression. Albert and Chib (1993) proposed a DA approach for sampling from the posterior distribution of the probit regression model. Their method makes use of auxiliary variables which allow an easy implementation of the Gibbs sampler. Polson et al. (2013) proposed an elegant algorithm for tackling the logistic regression case.

The generalized logistic distribution in equation (2) has mean zero and scale one. Consider now a generalized logistic with mean  $x_i^T \boldsymbol{\beta}$  and scale one, i.e. with probability density function:

$$f(x) = \frac{1}{B(p, p)} \frac{e^{p(x-x_i^T \boldsymbol{\beta})}}{(1 + e^{(x-x_i^T \boldsymbol{\beta})})^{2p}} \quad x \in \mathbb{R} \quad (5)$$

We denote the above distribution with  $\text{GLog}(x_i^T \boldsymbol{\beta}, 1)$ . Mimicking Albert and Chib (1993), suppose to sample  $Z_i$  from a  $\text{GLog}(x_i^T \boldsymbol{\beta}, 1)$ ,  $i = 1, \dots, n$ , and set  $Y_i = 1$  if  $Z_i > 0$ . Otherwise, if  $Z_i \leq 0$ , then set  $Y_i = 0$ . It is easy to see that

$$P(Y_i = 1) = \frac{1}{B(p, p)} B\left(\frac{e^{x_i^T \boldsymbol{\beta}}}{1 + e^{x_i^T \boldsymbol{\beta}}}; p, p\right).$$

The joint posterior distribution is

$$\pi(\boldsymbol{\beta}, p, \mathbf{Z} | \mathbf{y}) \propto \pi(\boldsymbol{\beta}) \pi(p) \prod_{i=1}^n \{\mathbb{I}(Z_i > 0) \mathbb{I}(y_i = 1) + \mathbb{I}(Z_i \leq 0) \mathbb{I}(y_i = 0)\}$$

$$\times \frac{1}{B(p, p)} \frac{e^{p(Z_i - x_i^T \boldsymbol{\beta})}}{(1 + e^{(Z_i - x_i^T \boldsymbol{\beta})})^{2p}} \quad (6)$$

where  $\pi(\boldsymbol{\beta})$  and  $\pi(p)$  are the prior distributions of  $\boldsymbol{\beta}$  and  $p$ , respectively. With the probit model, the technique of Albert and Chib (1993) works because the full conditional distributions can be resorted to normal distributions (or truncated normal distributions) which are easy to sample from. In our case, the usual normal prior would lead to a full conditional on  $\boldsymbol{\beta}$  which is not explicit, requiring the use of a Metropolis step (or alternative algorithms). The Pólya-Gamma identity in (1) allows to overcome this problem. In particular, we get

$$\frac{e^{p(Z_i - x_i^T \boldsymbol{\beta})}}{(1 + e^{(Z_i - x_i^T \boldsymbol{\beta})})^{2p}} = 2^{-2p} \int_0^{+\infty} e^{-\omega_i(Z_i - x_i^T \boldsymbol{\beta})^2/2} p(\omega_i) d\omega_i,$$

where  $p(\omega)$  is a Pólya-Gamma,  $\text{PG}(2p, 0)$ , distribution. Note that  $e^{-\omega(x_i^T \boldsymbol{\beta})^2/2} p(\omega)$  is the unnormalized density of a  $\text{PG}(2p, x_i^T \boldsymbol{\beta})$  random variable. Therefore, the augmented posterior distribution is

$$\begin{aligned} \pi(\boldsymbol{\beta}, p, \mathbf{Z}, \boldsymbol{\omega} | \mathbf{y}) &\propto \pi(\boldsymbol{\beta}) \pi(p) \prod_{i=1}^n \{ \mathbb{I}(Z_i > 0) \mathbb{I}(y_i = 1) + \mathbb{I}(Z_i \leq 0) \mathbb{I}(y_i = 0) \} \\ &\times 2^{-2p} e^{-\omega_i(Z_i - x_i^T \boldsymbol{\beta})^2/2} p(\omega_i). \end{aligned} \quad (7)$$

### 3 Inference Sampling Strategy

The algorithm proposed in Dalla Valle et al. (2019) makes use of an approximate likelihood, which relies on an approximation of the posterior distribution (see Mengersen et al. (2013) and Zhu and Leisen (2016)). The algorithm presented in this paper allows us to sample from the true posterior distribution, making it more appealing to perform the Bayesian inferential exercise.

Starting from the augmented posterior distribution in (7), this Section introduces the full conditional distributions required to run the algorithm. The variables involved are

$$\mathbf{Z}, \boldsymbol{\omega}, \boldsymbol{\beta}, p,$$

where  $\mathbf{Z} = (Z_1, \dots, Z_n)$ ,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$ .

**Full conditional for  $Z_i$ .** The random variables  $Z_1, \dots, Z_n$  are independent with full conditional distributions:

$$\begin{aligned} \pi(Z_i | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\omega}, p) &\propto \{\mathbb{I}(Z_i > 0) \mathbb{I}(y_i = 1) + \mathbb{I}(Z_i \leq 0) \mathbb{I}(y_i = 0)\} e^{-\omega_i(Z_i - x_i^T \boldsymbol{\beta})^2 / 2} \\ &\propto \{\mathbb{I}(Z_i > 0) \mathbb{I}(y_i = 1) + \mathbb{I}(Z_i \leq 0) \mathbb{I}(y_i = 0)\} e^{-\frac{\omega_i}{2}(Z_i^2 - 2Z_i x_i^T \boldsymbol{\beta})} \end{aligned}$$

which are truncated normal distributions. More precisely,

- if  $y_i = 1$ , then  $\pi(Z_i | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\omega}, p)$  is distributed as a  $\mathcal{N}(x_i^T \boldsymbol{\beta}, \omega_i^{-1})$  truncated to the left by 0;
- if  $y_i = 0$ , then  $\pi(Z_i | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\omega}, p)$  is distributed as a  $\mathcal{N}(x_i^T \boldsymbol{\beta}, \omega_i^{-1})$  truncated to the right by 0;

**Full conditional for  $\omega_i$ .** Following Polson et al. (2013), it is easy to see that the full conditional distribution of  $\omega_i$  is:

$$\pi(\omega_i | p, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{y}) \propto e^{-\omega_i(Z_i - x_i^T \boldsymbol{\beta})^2 / 2} p(\omega_i) \sim \text{PG}(2p, Z_i - x_i^T \boldsymbol{\beta})$$

An efficient method for sampling Pólya-Gamma random variables is implemented in a modified version of the R package `BayesLogit` (Windle et al., 2016).

**Full conditional for  $\boldsymbol{\beta}$ .** Assuming a multivariate  $\mathcal{N}(\boldsymbol{\beta}^*, B^*)$  prior for  $\boldsymbol{\beta}$ , the full conditional distribution of  $\boldsymbol{\beta}$  is

$$\pi(\boldsymbol{\beta} | \mathbf{Z}, \mathbf{y}, \boldsymbol{\omega}, p) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^n \exp \left\{ -\frac{\omega_i}{2} (Z_i - x_i^T \boldsymbol{\beta})^2 \right\}$$



$$\begin{aligned}
&= \exp \left\{ -\frac{1}{2} [\boldsymbol{\beta}^T (X^T \Omega X + B^{*-1}) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T (X^T \Omega Z + B^{*-1} \boldsymbol{\beta}^*)] \right\} \\
&\sim \mathcal{N}(\tilde{\boldsymbol{\beta}}^*, \tilde{V}_\beta^*)
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}^* &= (X^T \Omega X + B^{*-1})^{-1} (X^T \Omega Z + B^{*-1} \boldsymbol{\beta}^*) \\
\tilde{V}_\beta^* &= (X^T \Omega X + B^{*-1})^{-1}.
\end{aligned}$$

**Full conditional for  $p$ .** The full conditional for  $p$  is:

$$\pi(p | \boldsymbol{\beta}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{y}) \propto \pi(p) \prod_{i=1}^n \frac{2^{-2p}}{B(p, p)} p(\omega_i)$$

where  $p(\omega_i)$  is a  $\text{PG}(2p, 0)$ . In our approach we assume a Gamma prior for  $p$  and we integrate out the  $\omega$ 's in (7). This leads to the following full conditional distribution for  $p$ :

$$\begin{aligned}
\pi(p | \boldsymbol{\beta}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{y}) &\propto \pi(p) \frac{1}{B(p, p)^n} \prod_{i=1}^n \frac{e^{p(Z_i - x_i^T \boldsymbol{\beta})}}{(1 + e^{(Z_i - x_i^T \boldsymbol{\beta})})^{2p}} \\
&= \pi(p) \frac{1}{B(p, p)^n} \prod_{i=1}^n \left( \frac{e^{(Z_i - x_i^T \boldsymbol{\beta})}}{(1 + e^{(Z_i - x_i^T \boldsymbol{\beta})})^2} \right)^p.
\end{aligned}$$

In order to sample from  $p$ , we use the slice sampling of Neal (2003). This algorithm requires the computation of the inverse of the above full conditional distribution. We used the function `uniroot` available in R to compute the numerical inverse. The next sections illustrate the performance of the algorithm with simulated and real data.

## 4 Simulation Studies

In this section, we assess the performance of the algorithm by implementing different simulation experiments. In particular, we use four different values of the tail parameter  $p$ :  $p = 0.3$  and  $p = 0.7$  (heavy tails) and  $p = 1.5$  and  $p = 3$  (light tails). We focus on three scenarios:

1.  $\boldsymbol{\beta} = (0.5, -2, 1)$ , with  $k = 3$ ;
2.  $\boldsymbol{\beta} = (1, -1, -3, 1, 3)$ , with  $k = 5$ ;
3.  $\boldsymbol{\beta} = (2.3, 1, -2, 1.5, -2.7, 0.2, -1.4, 3, -0.6, -1.2)$ , with  $k = 10$ .

We assume a standard normal distribution for the explanatory variables. For each scenario and each value of  $p$ , we generate 20 different datasets of sample sizes  $n = 100$  and  $n = 250$ , respectively. Regarding the prior choice, we consider vague priors for  $\boldsymbol{\beta}$  and  $p$ . In particular, a multivariate normal distribution,  $\mathcal{N}(\boldsymbol{\nu}, B)$ , is chosen for  $\boldsymbol{\beta}$  with prior mean vector  $\boldsymbol{\nu} = 0$  and prior covariance matrix  $B = 5 \cdot I_k$ . For  $p$  we choose a gamma prior,  $\mathcal{G}a(a, b)$  with hyperparameters  $a = b = 1$ . For each of the 20 different simulated datasets, we run 3,000 iterations of the MCMC algorithm and discard the first 1,000 iterations as burn-in period.

In Table 1 we show the posterior means of the different values of  $p$  and  $n$  for the three-dimensional  $\boldsymbol{\beta}$  of scenario 1. The values in brackets show the standard deviations over the 20 simulations. We notice that the posterior means of  $p$  are very close to the true values, and increasing the dimensionality of  $n$  from 100 to 250 leads to a better estimation of the  $p$  parameter. This good performance is also confirmed for the posterior means of the vector of unknown coefficients  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ . We note an improvement of the results as the sample size increases.

Tables 2 and 3 show the posterior mean and the posterior standard deviations for the five- and ten-dimensional scenarios, respectively. These results are in line with the three-dimensional case, thus leading us to conclude that increasing the number of covariates is not an issue in terms of estimation accuracy. Moreover, the posterior standard deviations in Tables 1, 2 and 3 substantially decrease for all the different scenarios and  $p$  values when we move from  $n = 100$  to  $n = 250$ .

	$p$	$\beta_0$	$\beta_1$	$\beta_2$
<i>true values</i>	0.3	0.5	-2	1
$n = 100$	0.299 (0.002)	0.389 (0.467)	-1.772 (0.513)	0.926 (0.543)
$n = 250$	0.299 (0.001)	0.472 (0.263)	-1.929 (0.390)	1.092 (0.256)
<i>true values</i>	0.7	0.5	-2	1
$n = 100$	0.702 (0.005)	0.519 (0.249)	-2.163 (0.402)	1.138 (0.398)
$n = 250$	0.699 (0.003)	0.453 (0.203)	-2.018 (0.275)	1.012 (0.270)
<i>true values</i>	1.5	0.5	-2	1
$n = 100$	1.491 (0.014)	0.544 (0.283)	-2.129 (0.645)	1.016 (0.293)
$n = 250$	1.495 (0.007)	0.505 (0.163)	-2.078 (0.193)	1.109 (0.182)
<i>true values</i>	3	0.5	-2	1
$n = 100$	2.910 (0.033)	0.542 (0.169)	-2.114 (0.492)	1.095 (0.343)
$n = 250$	2.964 (0.014)	0.494 (0.094)	-2.079 (0.234)	1.028 (0.129)

Table 1: Posterior means over the 20 different simulated datasets for  $n$  equal to 100 and 250 compared with the true values of  $p = 0.3$  (first panel),  $p = 0.7$  (second),  $p = 1.5$  (third) and  $p = 3$  (forth) and  $(\beta_0, \beta_1, \beta_2) = (0.5, -2, 1)$ . The values in brackets are the standard deviations over the 20 different simulations.

	$p$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
<i>true values</i>	0.3	1	-1	-3	1	3
$n = 100$	0.304 (0.003)	0.872 (0.509)	-0.876 (0.495)	-2.877 (0.700)	0.853 (0.751)	2.720 (0.504)
$n = 250$	0.301 (0.001)	0.953 (0.331)	-1.058 (0.388)	-3.066 (0.405)	0.988 (0.419)	3.260 (0.427)
<i>true values</i>	0.7	1	-1	-3	1	3
$n = 100$	0.710 (0.011)	0.999 (0.351)	-0.988 (0.432)	-3.053 (0.602)	1.058 (0.401)	2.914 (0.343)
$n = 250$	0.704 (0.005)	0.947 (0.319)	-0.987 (0.355)	-3.053 (0.357)	0.961 (0.265)	3.142 (0.451)
<i>true values</i>	1.5	1	-1	-3	1	3
$n = 100$	1.523 (0.028)	0.988 (0.373)	-1.373 (0.459)	-3.422 (0.493)	1.127 (0.337)	3.233 (0.506)
$n = 250$	1.501 (0.013)	0.989 (0.194)	-1.067 (0.288)	-3.139 (0.334)	0.952 (0.202)	3.003 (0.337)
<i>true values</i>	3	1	-1	-3	1	3
$n = 100$	2.962 (0.053)	1.023 (0.234)	-1.182 (0.300)	-3.475 (0.623)	1.214 (0.428)	3.329 (0.597)
$n = 250$	2.990 (0.024)	1.048 (0.214)	-1.037 (0.201)	-3.161 (0.758)	1.080 (0.256)	3.173 (0.714)

Table 2: Posterior means over the 20 different simulated datasets for  $n$  equal to 100 and 250 compared with the true values of  $p = 0.3$  (first panel),  $p = 0.7$  (second),  $p = 1.5$  (third) and  $p = 3$  (forth) and  $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (1, -1, -3, 1, 3)$ . The values in brackets are the standard deviations over the 20 different simulations.



	$p$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$
<i>true values</i>	0.3	2.3	1	-2	1.5	-2.7	0.2	-1.4	3	-0.6	-1.2
$n = 100$	0.300 (0.007)	2.419 (0.631)	0.955 (0.607)	-1.926 (0.707)	1.147 (0.830)	-2.796 (0.698)	0.254 (0.664)	-1.214 (0.622)	2.982 (0.903)	-0.238 (0.689)	-0.906 (0.589)
$n = 250$	0.299 (0.002)	2.198 (0.564)	0.922 (0.513)	-1.985 (0.545)	1.527 (0.500)	-2.682 (0.583)	-0.024 (0.555)	-1.437 (0.351)	3.387 (0.650)	-0.594 (0.431)	-1.253 (0.456)
<i>true values</i>	0.7	2.3	1	-2	1.5	-2.7	0.2	-1.4	3	-0.6	-1.2
$n = 100$	0.701 (0.016)	2.333 (0.653)	1.070 (0.676)	-2.079 (0.645)	1.627 (0.536)	-2.814 (0.568)	0.187 (0.569)	-1.467 (0.488)	3.128 (0.837)	-0.684 (0.629)	-1.074 (0.650)
$n = 250$	0.700 (0.006)	2.437 (0.433)	1.085 (0.374)	-2.095 (0.423)	1.740 (0.455)	-2.918 (0.398)	0.234 (0.329)	-1.469 (0.344)	3.135 (0.567)	-0.625 (0.267)	-1.286 (0.306)
<i>true values</i>	1.5	2.3	1	-2	1.5	-2.7	0.2	-1.4	3	-0.6	-1.2
$n = 100$	1.502 (0.027)	2.475 (0.519)	1.290 (0.464)	-2.090 (0.585)	1.545 (0.424)	-2.874 (0.568)	0.346 (0.447)	-1.501 (0.569)	3.144 (0.718)	-0.587 (0.417)	-1.160 (0.340)
$n = 250$	1.502 (0.020)	2.518 (0.471)	1.077 (0.243)	-2.134 (0.441)	1.647 (0.307)	-2.916 (0.559)	0.185 (0.284)	-1.498 (0.360)	3.379 (0.566)	-0.658 (0.245)	-1.374 (0.400)
<i>true values</i>	3	2.3	1	-2	1.5	-2.7	0.2	-1.4	3	-0.6	-1.2
$n = 100$	2.933 (0.065)	2.713 (0.590)	1.159 (0.509)	-2.295 (0.546)	1.788 (0.494)	-3.114 (0.661)	0.286 (0.534)	-1.867 (0.756)	3.228 (0.679)	-0.741 (0.386)	-1.345 (0.517)
$n = 250$	2.972 (0.036)	2.462 (0.305)	1.080 (0.257)	-2.089 (0.327)	1.601 (0.309)	-2.901 (0.398)	0.147 (0.202)	-1.561 (0.211)	3.247 (0.466)	-0.669 (0.229)	-1.253 (0.259)

Table 3: Posterior means over the 20 different simulated datasets for  $n$  equal to 100 and 250 compared with the true values of  $p = 0.3$  (first panel),  $p = 0.7$  (second),  $p = 1.5$  (third) and  $p = 3$  (forth) and  $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9) = (2.3, 1, -2, 1.5, -2.7, 0.2, -1.4, 3, -0.6, -1.2)$ . The values in brackets are the standard deviations over the 20 different simulations.

In order to illustrate the complete inferential procedure, we show how all the parameters of the model are estimated. Figure 2 displays the posterior chains for a specific dataset with sample size  $n = 250$ , where  $p = 0.3$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (1, -1, -3, 1, 3)$ . As shown in Figure 2, the chains, after a burn-in of 1,000 iterations, converge quickly to the true values of all parameters and give an excellent representation of the real parameter values.

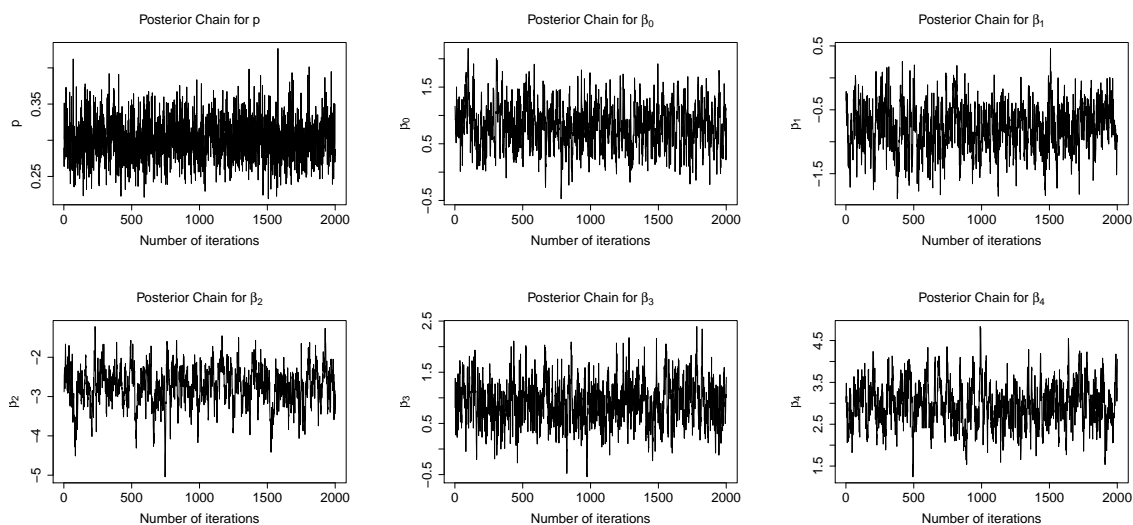


Figure 2: Sample chains of the posterior distribution of the parameters for the simulated data with sample size  $n = 250$  and with true parameter values  $p = 0.3$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (1, -1, -3, 1, 3)$ .

We also performed a convergence analysis for scenario 1, with three-dimensional  $\boldsymbol{\beta}^1$ . The analysis was carried out using the R `coda` package (Plummer et al., 2006). In particular, we used single datasets of sample sizes  $n = 100$  and  $n = 250$ , respectively, where we fixed  $p = 0.3$ . We computed Geweke’s convergence tests and the autocorrelation and partial autocorrelation

---

<sup>1</sup>We performed other convergence tests for the five- and ten-dimensional cases. These results are omitted for lack of space but are available on request.

functions for the values of  $p$  and  $\boldsymbol{\beta}$ . Table 4 reports the results of Geweke’s convergence test (Geweke, 1992), that indicate no issues since the test statistics values are all lower than 1.96. As a final check, in Figures 3 and 4 we plotted the autocorrelation and the partial autocorrelation functions for the three-dimensional scenario with  $n = 100$  and  $n = 250$ , respectively. From these Figures, there is no indication of serious convergence problems.

	Variables	Test Statistic		Variables	Test Statistic
$n = 100$	$p = 0.3$	0.604	$n = 250$	$p = 0.3$	0.8352
	$\beta_0 = 0.5$	0.7215		$\beta_0 = 0.5$	-0.2248
	$\beta_1 = -2$	-0.3355		$\beta_1 = -2$	0.6250
	$\beta_2 = 1$	1.4029		$\beta_2 = 1$	-0.6288

Table 4: Geweke’s test statistics for the posterior chain of the tail parameter  $p$  and of the unknown coefficients,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$  for a simulated dataset of sample size  $n = 100$  (left panel) and  $n = 250$  (right panel).

## 5 Real Data Applications

In this Section, we present two different real data applications to illustrate the accurate and efficient performance of the Pólya-Gamma sampler for estimating the generalized logistic regression. The applications analyse novel datasets of different dimensions, which are related to current and topical problems in the social sciences. In the first application, we consider data selected from the European Social Survey (ESS), that were previously analysed in Dalla Valle et al. (2019) to identify the determinants of public opinions towards immigration. The second application analyses the motives behind the recidivism of offenders released from prison between



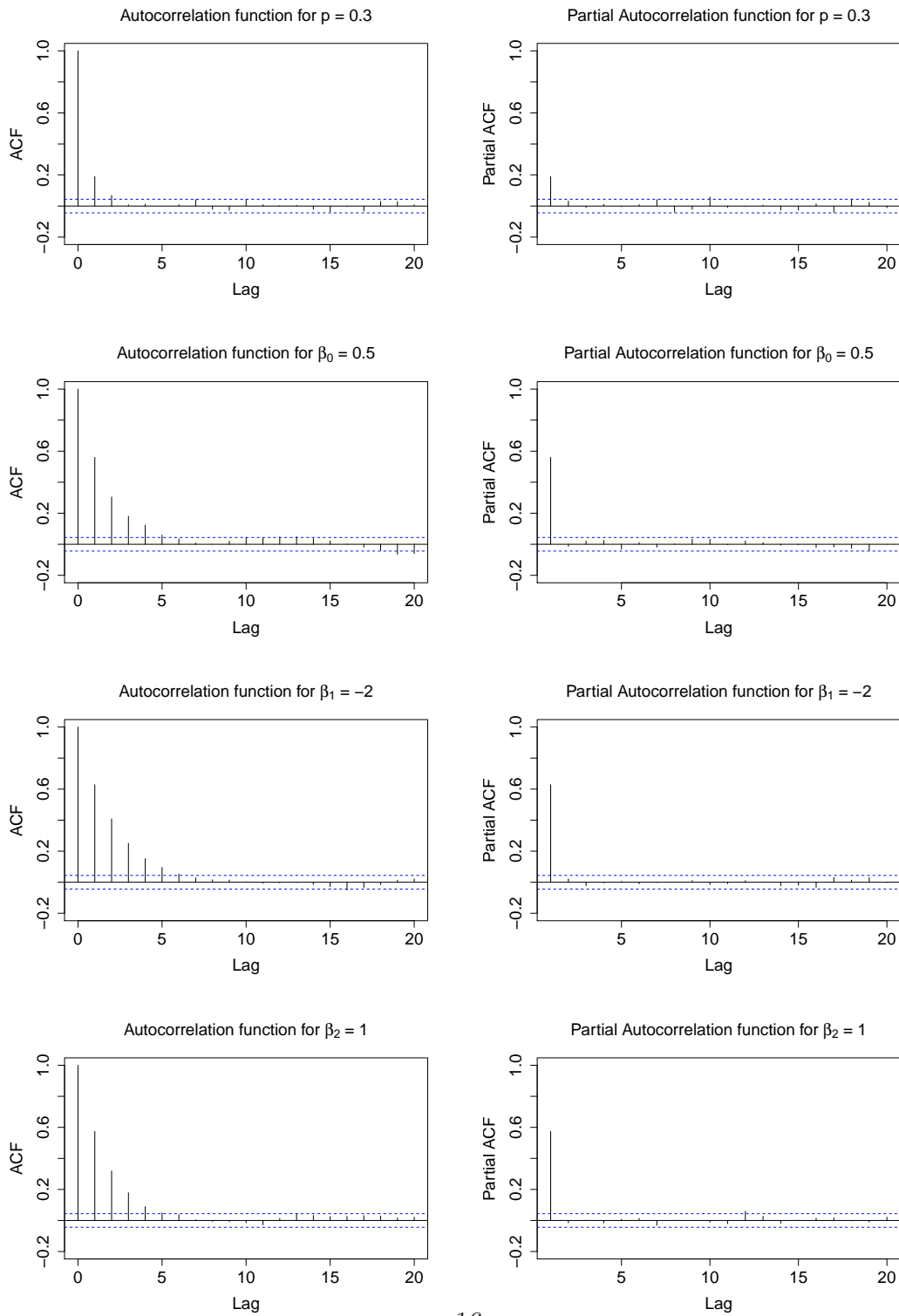


Figure 3: Autocorrelation (left) and partial autocorrelation (right) functions for the simulated dataset of sample size  $n = 100$ , with true parameter values  $p = 0.3$  (first row) and  $\beta = (\beta_0, \beta_1, \beta_2) = (0.5, -2, 1)$  (second, third and fourth row).

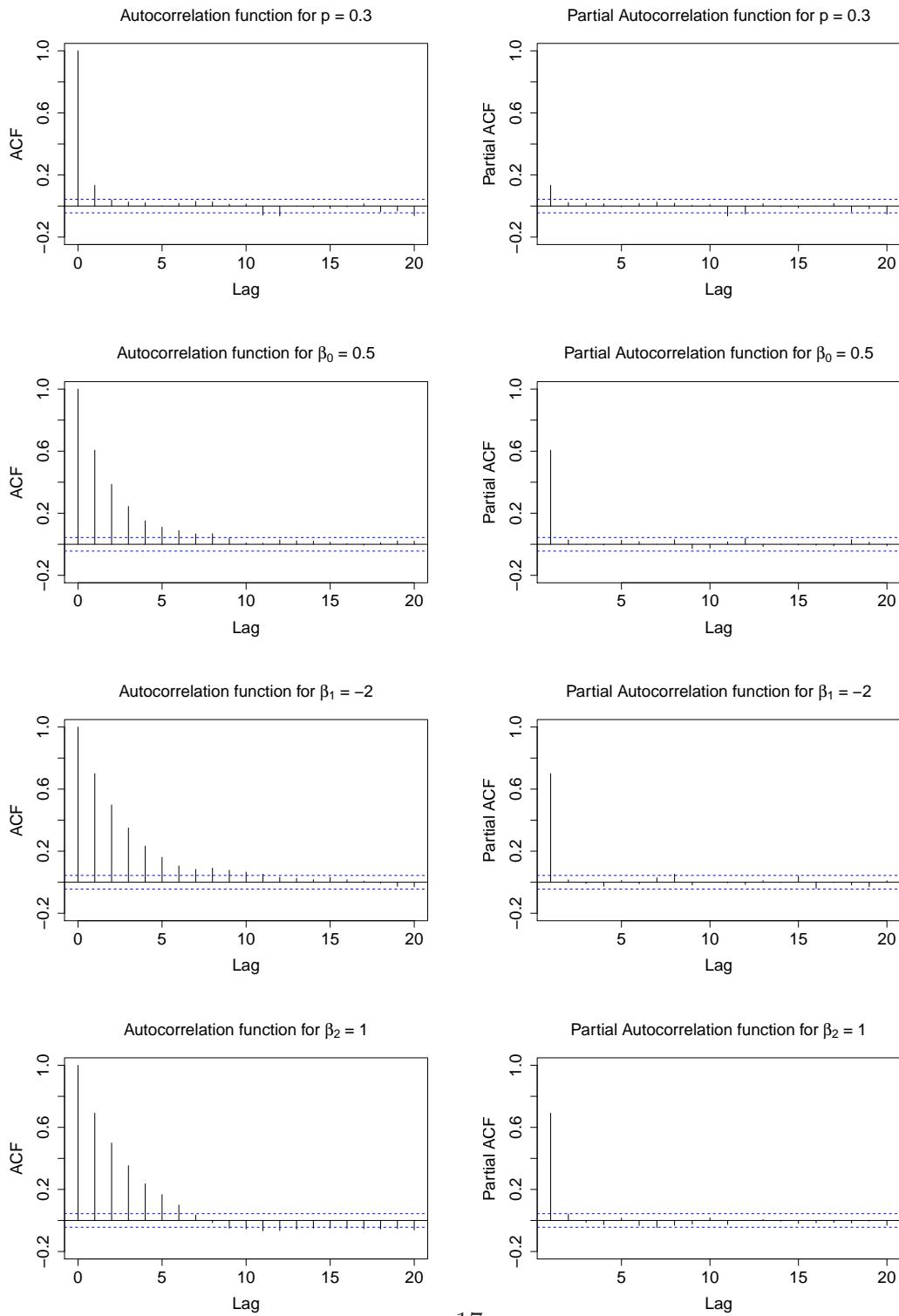


Figure 4: Autocorrelation (left) and partial autocorrelation (right) functions for the simulated dataset of sample size  $n = 250$ , with true parameter values  $p = 0.3$  (first row) and  $\beta = (\beta_0, \beta_1, \beta_2) = (0.5, -2, 1)$  (second, third and fourth row).

2016 and 2018 in the State of Iowa, USA.

## 5.1 Immigration in European Countries

The first dataset focuses on people’s attitudes towards immigration. The data was selected from the ESS and was collected following hour-long face-to-face interviews. We focused on questions regarding immigration from the ESS8 edition 1.0 published in October 2017 and collected between August 2016 and March 2017 (ESS8, 2016, 2018) for Great Britain (GB), Germany (DE) and France (FR), respectively. For these three countries, the total number of observations is 5,354, of which 1,419 are for GB; 2,284 for DE and 1,651 for FR.

The dependent variable is `immig`, that indicates whether the respondent would allow immigrants from poorer countries outside Europe. More precisely, if `immig` = 1 the respondent is against immigration, if `immig` = 0 the respondent is in favor of immigration. Following Dalla Valle et al. (2019), we considered as covariates the variables: `pplfair`, `trstep`, `trstun`, `happy`, `agea`, `edulvlb` and `hinctnta`. The first three variables include answers to question ranging from 0 (most negative) to 10 (most positive): do you think that most people try to take advantage of you (`pplfair`)? Do you trust the European parliament (`trstep`)? Do you trust the United Nations (`trstun`)? The `happy` variable takes values from 0 (unhappy) to 10 (extremely happy). The last three variables are subject-specific and include the age, ranging from 15 to 100 (`agea`); the highest level of education, from primary education to doctoral degree (`edulvlb`) and the household’s total net income (`hinctnta`).

For this dataset, we run 3,000 iterations of the Pólya-Gamma sampler algorithm and we discarded the first 1,000 iterations as burn-in. We used the MLE estimates of  $\beta$  and  $p$  as starting points of the algorithm and we employed vague priors, as in the simulated examples.

In order to assess the performance of the Pólya-Gamma sampler, we report the results obtained using the empirical likelihood approach on the same dataset, adopting a similar setting.

Table 5 lists the results of the Pólya-Gamma sampler (left-hand side) and the empirical likelihood approach (right-hand side). For each European country (GB, DE and FR) and for each parameter, Table 5 shows the posterior means, the 2.5% and 97.5% quantiles of the posterior distributions, the credible interval (CI) widths and the difference between the CIs of the Pólya-Gamma and the empirical likelihood methods. The effects of the parameters are similar across countries and suggest that the higher the level of education, the higher the probability for people to be favourable towards immigration. Also, trustful people and individuals with a strong confidence in the European parliament tend to be in favour of immigration. The posterior mean estimates obtained with the Pólya-Gamma method are in line with those obtained with the empirical likelihood. However, the CIs calculated with the first approach are narrower than those calculated with the second approach. This is illustrated in the last column of Table 5, where the negative values show that the Pólya-Gamma CI widths are smaller than those of the empirical likelihood method. For some of the covariates, such as `pplfair`, `trstep`, `agea` and `hinctnta`, the empirical likelihood approach assigns posterior support to the value zero, while the corresponding Pólya-Gamma CIs do not include zero. In addition, although the estimates of the  $p$  parameter for all countries suggest heavy tails for the generalised logistic regression, the empirical likelihood CIs cover the value 1, which correspond to the standard logistic regression. On the contrary, the results of the Pólya-Gamma method give no posterior support to the value 1 for DE and FR data, and very limited support for BG data, clearly suggesting the need for the generalised logistic regression model. Therefore, the Pólya-Gamma sampler outperforms the empirical likelihood method for the estimation of the generalised logistic regression, since it provides more precise parameter estimates as for the  $\beta$  as well as for the  $p$  parameters.

		Pólya-Gamma Sampler				Empirical Likelihood				
Country	Parameters	post. mean	2.50%	97.50%	CI width	post. mean	2.50%	97.50%	CI width	CI difference
GB	const	-0.6437	-1.6623	0.294	1.9563	-0.4202	-2.4377	1.4487	3.8864	-1.9301
	pplfair	-0.1463	-0.233	-0.0567	0.1763	-0.1822	-0.3176	0.0113	0.3289	-0.1526
	trstep	-0.2265	-0.3275	-0.1311	0.1964	-0.1731	-0.3796	0.0447	0.4243	-0.2279
	trstun	-0.0423	-0.1282	0.0407	0.1689	-0.0372	-0.2354	0.1409	0.3763	-0.2074
	happy	0.0343	-0.0581	0.1269	0.185	-0.0133	-0.2892	0.1209	0.4101	-0.2251
	agea	0.009	-0.002	0.0206	0.0226	0.0094	-0.0139	0.0358	0.0497	-0.0271
	edulvlb	-0.0023	-0.0034	-0.0011	0.0023	-0.0021	-0.0039	-0.0006	0.0033	-0.001
	hinctnta	0.012	-0.0613	0.0892	0.1505	-0.044	-0.1856	0.1228	0.3084	-0.1579
	$p$	0.928	0.8508	1.008	0.1572	0.8218	0.4284	1.5696	1.1412	-0.984
DE	const	-0.9217	-1.9118	0.0194	1.9312	-1.1452	-2.4617	0.6941	3.1558	-1.2246
	pplfair	-0.1683	-0.2549	-0.0721	0.1828	-0.1612	-0.3291	-0.0093	0.3198	-0.137
	trstep	-0.211	-0.3203	-0.1095	0.2108	-0.1524	-0.2839	-0.0107	0.2732	-0.0624
	trstun	-0.0441	-0.1488	0.0535	0.2023	-0.0112	-0.2115	0.1847	0.3962	-0.1939
	happy	-0.0796	-0.1767	0.0174	0.1941	-0.0873	-0.2438	0.0659	0.3097	-0.1156
	agea	0.0194	0.0081	0.0305	0.0224	0.0124	-0.0061	0.0375	0.0436	-0.0212
	edulvlb	-0.0001	-0.0013	0.0014	0.0027	-0.0007	-0.0026	0.0016	0.0042	-0.0015
	hinctnta	-0.1009	-0.1857	-0.0189	0.1668	-0.1171	-0.2352	0.0414	0.2766	-0.1098
	$p$	0.7903	0.739	0.843	0.104	0.7338	0.2955	1.7036	1.4081	-1.3041
FR	const	0.692	-0.3783	1.7869	2.1652	0.8442	-0.3443	1.4314	1.7757	0.3895
	pplfair	-0.0824	-0.1648	0.004	0.1688	-0.0921	-0.1898	0.0193	0.2091	-0.0403
	trstep	-0.2489	-0.3561	-0.1394	0.2167	-0.1912	-0.3592	-0.0114	0.3478	-0.1311
	trstun	-0.0895	-0.1829	0.0042	0.1871	-0.0514	-0.2047	0.1417	0.3464	-0.1593
	happy	0.029	-0.0673	0.1256	0.1929	0.0191	-0.0609	0.1545	0.2154	-0.0225
	agea	-0.0006	-0.0111	0.0103	0.0214	-0.0043	-0.0155	0.0091	0.0246	-0.0032
	edulvlb	-0.0037	-0.0052	-0.0024	0.0028	-0.0038	-0.0056	-0.0027	0.0029	-0.0001
	hinctnta	-0.0666	-0.1403	0.013	0.1533	-0.1061	-0.239	0.0286	0.2676	-0.1143
	$p$	0.7317	0.6745	0.7905	0.116	0.7845	0.3991	1.2809	0.8818	-0.7658

Table 5: EU immigration data results for Great Britain (GB), Germany (DE) and France (FR), obtained with the Pólya-Gamma (left-hand side) and the empirical likelihood (right-hand side) methods. For each parameter, columns 3 and 7 show the posterior means, columns 4 and 8 show the 2.5% quantiles of the posterior distribution, columns 5 and 9 show the 97.5% quantiles of the posterior distribution, columns 6 and 10 show the credible interval (CI) widths and column 11 shows the difference between the CIs of the Pólya-Gamma and the empirical likelihood methods.

## 5.2 Recidivism of Offenders Released from Prison in Iowa

The second dataset contains information about the recidivism of offenders released from prison between 2016 and 2018 in Iowa<sup>2</sup>. Recidivism happens when a prisoner is released from jail and he or she relapses back into criminal behavior and returns to jail. We applied the generalised logistic regression to determine whether a prisoner is likely to recidivate based on his or her characteristics.

After removing missing values, the data consists of 13,644 records corresponding to offenders detained in prison and released from 2016 to 2018. The dependent variable is `recidivism`, which is equal to 1 if the prisoner is arrested again within a 3-year tracking period after being released, and it is equal to 0 if he or she has not returned to prison within the tracking period.

We model the probability of being recidivist using a subsample of all the explanatory variables present in the original database: the sex of the offender (`sex`, which is equal to 1 for men and 0 for women); the age when released from prison (`age`); whether an offender committed a felony or misdemeanor (`felony`, which is equal to 1 for felony and 0 for misdemeanor); whether an offender was charged with a drug crime (`drug`); whether an offender was charged with a public order crime (`puborder`); whether an offender was charged with a violent crime (`violent`); and whether an offender was released on discharge or parole (`discharge`, which is equal to 1 for discharge and 0 for parole).

In this second example, we run the MCMC algorithm using 2,000 iterations and discarding the first 500 iterations. As in the immigration example, we adopted vague priors and we stated the initial values of  $\beta$  and  $p$  based on the corresponding MLE estimators. Figure 5 shows the posterior chain of the tail parameter  $p$ , which demonstrates a good convergence performance.

As we did with the immigration data, we compared the performances of the Pólya-Gamma

---

<sup>2</sup>See the following website for details: <https://data.iowa.gov/Correctional-System/3-Year-Recidivism-for-Offenders-Released>

Parameters	Pólya-Gamma Sampler				Empirical Likelihood				CI difference
	post. mean	2.50%	97.50%	CI width	post. mean	2.50%	97.50%	CI width	
const	-0.1096	-0.2295	0.0154	0.2449	-0.4278	-0.9341	0.1719	1.1060	-0.8611
sex	0.2252	0.1349	0.3154	0.4503	0.1371	-0.2583	0.4574	0.7157	-0.2654
age	-0.0002	-0.0003	-0.0002	0.0001	-0.0003	-0.0005	-0.0002	0.0003	-0.0002
felony	-0.0678	-0.1471	0.0137	0.1608	-0.0974	-0.4457	0.0999	0.5456	-0.3842
drug	-0.0268	-0.1020	0.0464	0.1484	-0.0996	-0.4009	0.1912	0.5921	-0.4437
puborder	-0.2000	-0.2949	-0.0998	0.1951	-0.0694	-0.3384	0.2261	0.5645	-0.3694
violent	-0.4118	-0.4958	-0.3204	0.1754	-0.4743	-0.9779	-0.2638	0.7141	-0.5387
discharge	-0.4085	-0.4959	-0.3209	0.1750	-0.4654	-0.6726	-0.1936	0.4790	-0.3040
$p$	1.3048	1.2656	1.3450	0.0794	0.5679	0.3409	1.0905	0.7496	-0.6702

Table 6: Criminal recidivism data results obtained with the Pólya-Gamma (left-hand side) and the empirical likelihood (right-hand side) methods. For each parameter, columns 2 and 6 show the posterior means, columns 3 and 7 show the 2.5% quantiles of the posterior distribution, columns 4 and 8 show the 97.5% quantiles of the posterior distribution, columns 5 and 9 show the credible interval (CI) widths and column 10 shows the difference between the CIs of the Pólya-Gamma and the empirical likelihood methods.

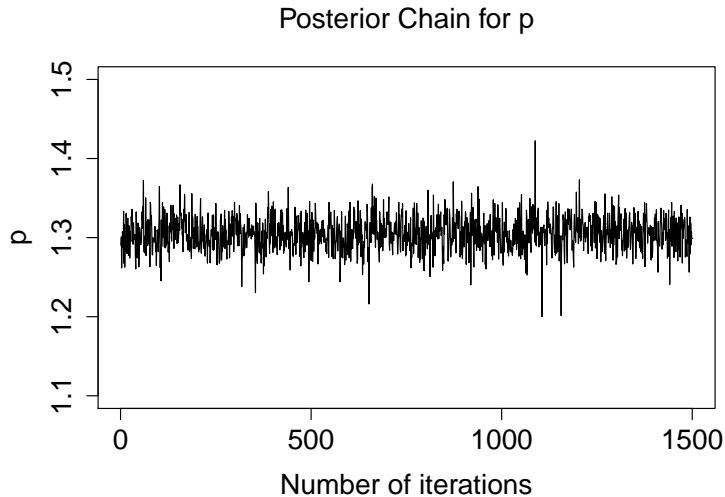


Figure 5: Posterior chain for the tail parameter  $p$  for the recidivism dataset.

sampler and the empirical likelihood approaches, using the same settings for both algorithms. The results for the parameters  $\beta$  and  $p$  are listed in Table 6, that displays the Pólya-Gamma results on the left-hand side and the empirical likelihood results on the right-hand side. Table 6 illustrates the posterior means, the 2.5% and 97.5% quantiles of the posterior distributions, the credible interval (CI) widths and the difference between the CIs of the Pólya-Gamma and the empirical likelihood methods.

The results suggest that men are more likely to recidivate than women, younger prisoners are more likely to become repeat offenders than older prisoners, there is a lower probability that the criminal is a recidivist if he or she was charged with a public order or violent crime and offenders released on parole are more likely to return to jail than those released on discharge.

The sign of the posterior means for both approaches is the same, suggesting a similar direction for the effects obtained using the Pólya-Gamma and the empirical likelihood method. However, the CIs of the first approach are again narrower than those of second approach, as



shown by the CIs differences in the last column of Table 6 , which are all negative. Moreover, the Pólya-Gamma method, unlike the empirical likelihood, does not give posterior support to zero to the variables `sex` and `puborder`, suggesting that the inclusion of these variables in the model is appropriate.

The posterior estimates of  $p$  are different for the two approaches. The posterior mean estimated with the Pólya-Gamma method is greater than one, denoting light tails for the generalised logistic model, and there is no posterior support for the value one, suggesting that the tails need to be appropriately modelled. On the contrary, the posterior mean estimated with the empirical likelihood is lower than one and the CI is wide and includes the values one, denoting estimation inaccuracy and uncertainty.

In agreement with our findings with the immigration data, the results obtained with the recidivism dataset demonstrate that the Pólya-Gamma sampler method yields a higher estimates precision and accuracy compared to the empirical likelihood method, as for the covariate parameters  $\beta$  as well as for the tail parameter  $p$ . In particular, a more precise estimation of  $p$  leads to more clarity on the tail behaviour of the generalised logistic regression.

## 6 Conclusions

This paper introduces a novel DA scheme for the generalized logistic regression model. This model is particularly flexible since it is able to accommodate both light and heavy tails in dichotomous response data. The proposed DA scheme makes use of the Pólya-Gamma identity and it is strongly related to the slice sampler algorithm. The Pólya-Gamma sampler allows us to implement a Bayesian method able to draw samples from the exact posterior distribution. On the contrary, other methods, such as the empirical likelihood approach, are based on approximations of the posterior and may lead to low precision in the estimation of the parameters. Our

simulation study demonstrates the estimation accuracy of the newly proposed Pólya-Gamma sampler with datasets of different dimensions. We compared the performances of the Pólya-Gamma and the empirical likelihood methods for modelling new interesting datasets regarding the opinion on immigration in European countries and the probability of being recidivist for prisoners in Iowa, USA. Our results demonstrate the superiority of the Pólya-Gamma sampler over the empirical likelihood in terms of parameter precision. The Pólya-Gamma method allows a more accurate estimation of the tail parameter for the generalised logistic regression compared to approximate methods.

## References

- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.
- Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). Handbook of markov chain monte carlo.
- Choi, H. M. and Hobert, J. P. (2013). The poly-gamma gibbs sampler for bayesian logistic regression is uniformly ergodic. *Electron. J. Statist.*, 7:2054–2064.
- Dalla Valle, L., Leisen, F., Rossini, L., and Zhu, W. (2019). Bayesian analysis of immigration in europe with generalized logistic regression. *Journal of Applied Statistics*, 0(0):1–15.
- ESS8 (2016). ESS Round 8: European Social Survey Round 8 Data. *Data file edition 2.0. NSD - Norwegian Centre for Research Data, Norway ??? Data Archive and distributor of ESS data for ESS ERIC.*
- ESS8 (2018). ESS Round 8: European Social Survey. *ESS-8 2016 Documentation Report.*

*Edition 2.0. Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC.*

- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, J. F. M., editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford.
- Hensley, A. A. and Djurić, P. M. (2017). Nonparametric learning for hidden markov models with preferential attachment dynamics. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3854–3858.
- Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and px-da algorithms. *Ann. Statist.*, 36(2):532–554.
- Holmes, C. C., Held, L., et al. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1):145–168.
- Karabatsos, G. and Leisen, F. (2018). An approximate likelihood perspective on abc methods. *Statist. Surv.*, 12:66–104.
- Leisen, F., Rossini, L., and Villa, C. (2017). A note on the posterior inference for the yule–simon distribution. *Journal of Statistical Computation and Simulation*, 87(6):1179–1188.
- Liu, J. and Wu, Y. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94.
- Meng, X.-L. and van Dyk, D. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, 86(2):301–320.
- Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, 110(4):1321–1326.

- Neal, R. M. (2003). Slice sampling. *Ann. Statist.*, 31(3):705–767.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.
- Tanner, T. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.
- van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50.
- Windle, J., Polson, N. G., and Scott, J. G. (2016). BayesLogit: Bayesian Logistic Regression. Technical report, R Package Version 0.6.
- Zhu, W. and Leisen, F. (2016). A bootstrap likelihood approach to bayesian computation. *Australian and New Zealand Journal of Statistics*, 58(1):227–244.