

## Introduction

Modelling binary data implies the choice of link function and the choice of covariate set. The covariate set uncertainty has been explored in the literature, mainly through Bayesian Model Averaging (BMA). In the case of link function uncertainty less research has been done. We propose an algorithm that considers both uncertainties within a Bayesian framework. The regression model for binary data is an extension to the Albert and Chib (1993) model. It is defined as a scale mixture of normals linear model based on latent continuous data to deal with link function uncertainty. This approach allows for flexible shapes which do not necessarily correspond to a known model. The set of covariates that are used in the model is also left free and estimated from the data using the Green (1995) RJMCMC algorithm.

## Link Function Uncertainty

When modelling binary data the link functions commonly used in practice are the Probit, or inverse Normal, and the Logistic: the first because of its relation with the normal distribution, the second due to its interpretation with the Log-Odds ratio. Nevertheless, there is no specific justification for using one or the other. In practice, we often choose one, and do not consider the other case. However, this selection fails to take into account the prior uncertainty in the link function.

Albert and Chib (1993), propose the use of auxiliary variables and a scale mixture of normals to represent the noise process in Bayesian binary regression.

The model, defined in terms of the auxiliary variable  $z_i$ , is as follows:

For  $y_i \in \{0, 1\}$ ,  $i = 1, \dots, n$  a binary variable for data size  $n$  and covariate vector  $x_i = (x_{i1}, \dots, x_{ip})$  of dimension  $p$ .

$$y_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $z_i = x_i\beta + \varepsilon_i$  and  $\varepsilon_i \sim N(0, \lambda_i^{-1})$ . With  $\beta$  a vector of regression coefficients,  $z_i$  is a linear predictor with noise  $\varepsilon_i$  with mean zero and scale determined by the distribution of  $\lambda_i \sim f(\cdot)$ .

Therefore the binary model is defined as:  $Y_i \sim \text{Ber}(p_i)$  with  $p_i = \Phi(x_i\beta\sqrt{\lambda_i})$  the cumulative standard normal distribution evaluated at  $x_i\beta\sqrt{\lambda_i}$ .

The binary data is modeled as a truncated normal, and so **the conditional distribution given  $\lambda$  is equivalent to those under Bayesian linear regression model with Gaussian noise**. Then the posterior distribution of the parameters  $\beta$  can be computed using standard results for normal linear models. By using various mixing distributions  $f(\cdot)$  for  $\lambda_i$ , we allow flexible shapes, as heavier tails, for the augmented data  $z_i$ . Therefore the distribution of  $Y_i$  depends on the distribution of the scale parameter  $\lambda_i$ .

## Choices for $\lambda_i$ :

1.  $\lambda_i = 1$ . This is the particular case where  $\lambda_i$  is a constant, and generates the known **Probit** model.
2.  $\lambda_i = (2\psi_i)^{-2}$  with  $\psi_i$  distributed as a Kolmogorov-Smirnov. The scale mixture distribution generates the known **Logit** model. See Holmes and Held (2006).
3.  $\lambda_i \sim \text{Ga}(\nu/2, \nu/2)$ . This distribution is well known in literature, it generates a **Student- $t_\nu$**  distribution for  $\varepsilon_i$ .
  - In the limit as  $\nu \rightarrow \infty$ :  $\varepsilon_i \rightarrow N(0, 1)$  and therefore the binary model  $\mathcal{M}$  approximates a **Probit**.
4.  $\lambda_i \sim \text{LogN}(-\nu/2, \nu)$ . Our proposal for the mixing distribution.
  - In the limit, as  $\nu \rightarrow 0$  the scale mixture of normals tends to a  $N(0, 1)$  and the binary model approximates a **Probit**.

The link function is defined through the scale mixture of normals by the distribution of  $\lambda_i$ . We adopt a fully Bayesian approach by letting the mixing parameters  $\nu$  and  $v$  of  $\lambda_i$  be estimated from the data. We propose two models that consider uncertainty about the link function, and allow for inference on the shape of the link function from the data. Therefore, within the context of link function uncertainty  $\nu$  and  $v$  will be the model determinant parameters.

## Extension to covariate set uncertainty

We consider model uncertainty for covariate selection by defining a covariate indicator vector  $\gamma = \{\gamma_1, \dots, \gamma_p\}$ , with  $p$  the total number of plausible covariates. Here  $\gamma_i = 1$  indicates that the covariate  $i$  is included in the model and  $\gamma_i = 0$  indicates it is not, for all  $i = 1, \dots, p$ . This vector is part of the model and its update provides the posterior model probability:  $\pi(\gamma|D)$ .

We increase a hierarchy in the scale mixture of normals model to incorporate the parameter vector  $\gamma$  and follow Holmes and Held (2006) approach for jointly updating  $\{z, \beta\}$ .

## MCMC algorithm

We proceed with a Hybrid Gibbs sampler, as suggested by Holmes and Held (2006), the model and parameter coefficient vector  $\{\gamma, \beta\}$  are updated together using the Reversible Jump Monte Carlo Markov Chain (RJMCMC), as proposed by Green (1995). Once the model  $\gamma$  is determined, we consider Fernandez and Steel (2000) sampling proposal for the update of mixing distribution. This includes a Metropolis-Hasting step for the update of  $\nu$  and  $v$ , and rejection sampling for the update of  $\lambda_i$  in the Lognormal mixture.

## Application to the Pima Indian Data

We apply the sampling algorithms to the Pima Indian Data, referred to in Holmes and Held (2006). The data consists of seven plausible covariates and 532 observations to model the outcome of a diabetes test. The covariates are: number of pregnancies (NP), plasma glucose concentration (GI), diastolic blood pressure (BP), triceps skin fold thickness (TST), body mass index (BMI), diabetes pedigree function (DP), and age (AG).

To illustrate the method we perform a simulation exercise, following Holmes and Held (2006). We compare their results, *Logit (HH)*, with the ones we obtain using the following models and methods: *Logit (MH)* refers to the Logit Metropolis-Hasting algorithm. *Probit* is the Probit modelled using the augmented data approach and iterative method algorithm. *Mixture Gamma* and *Mixture LogNormal* are the scale mixture of normals proposals, using the joint update sampling algorithm for  $\{\beta, \gamma\}$  and iterative sampling for  $\{z, \lambda, \theta\}$ .

In order to be comparable with each other and with Holmes and Held's (2006) results, we run each algorithm for 10,000 iterations, after a burn-in period of 1,000, all starting with the same random seed to avoid other sources of uncertainty. All algorithms start with the full model  $\gamma = 1111111$ . We use the same prior  $\pi(\beta) = N(0, 100I_p)$  in all models.

Figure 1 compares Holmes and Held's (2006) estimate of the posterior probability  $E[\gamma_j] = \pi(\gamma_j = 1|y)$  to the ones we obtain from the simulation exercise.

MODEL	NP	GI	BP	TST	BMI	DP	Ag
Logit (HH)	0.923	0.999	0.009	0.037	0.993	0.944	0.129
Logit (MH)	0.9862	1	0.0422	0.0906	1	0.9916	0.2669
Probit	0.9978	1	0.0016	0.005	0.9017	0.5746	0.0006
Gamma Mixture	0.9957	1	0.0014	0.002	0.8701	0.5084	0.0051
LogNormal Mixture	0.9585	1	0.0021	0.0217	0.8815	0.724	0.036

Figure 1: Posterior probability of including  $\beta_j$

In general, the estimates of  $E[\gamma_j]$  in our proposal are similar to the ones reported by Holmes and Held. However, except for the *Logit (MH)*, we observe some differences in the diabetes pedigree (DP) and the Age (AG). The largest difference is observed in the DP regressor. Therefore **it is suggested, by the difference between Logit, Probit, Gamma and LogNormal mixture models, that the inclusion of a covariate is related to the choice of link function**.

The difference in models can also be appreciated in Figure 2, where for each percentile of the posterior model probability (ordered by more likely) we plot the binary vector  $\gamma$ . The inclusion of a covariate is shaded in grey, and the exclusion in black. Covariates 1, 2 and 5 (NP, GI and BMI) are included in virtually all models. For the 6th covariate (DP), we notice the difference between models. Notice that even though the posterior probability of including it in the models Probit and Gamma Mixture is very similar, the order is not the same. In the case of considering only the most likely model we would have had different results.

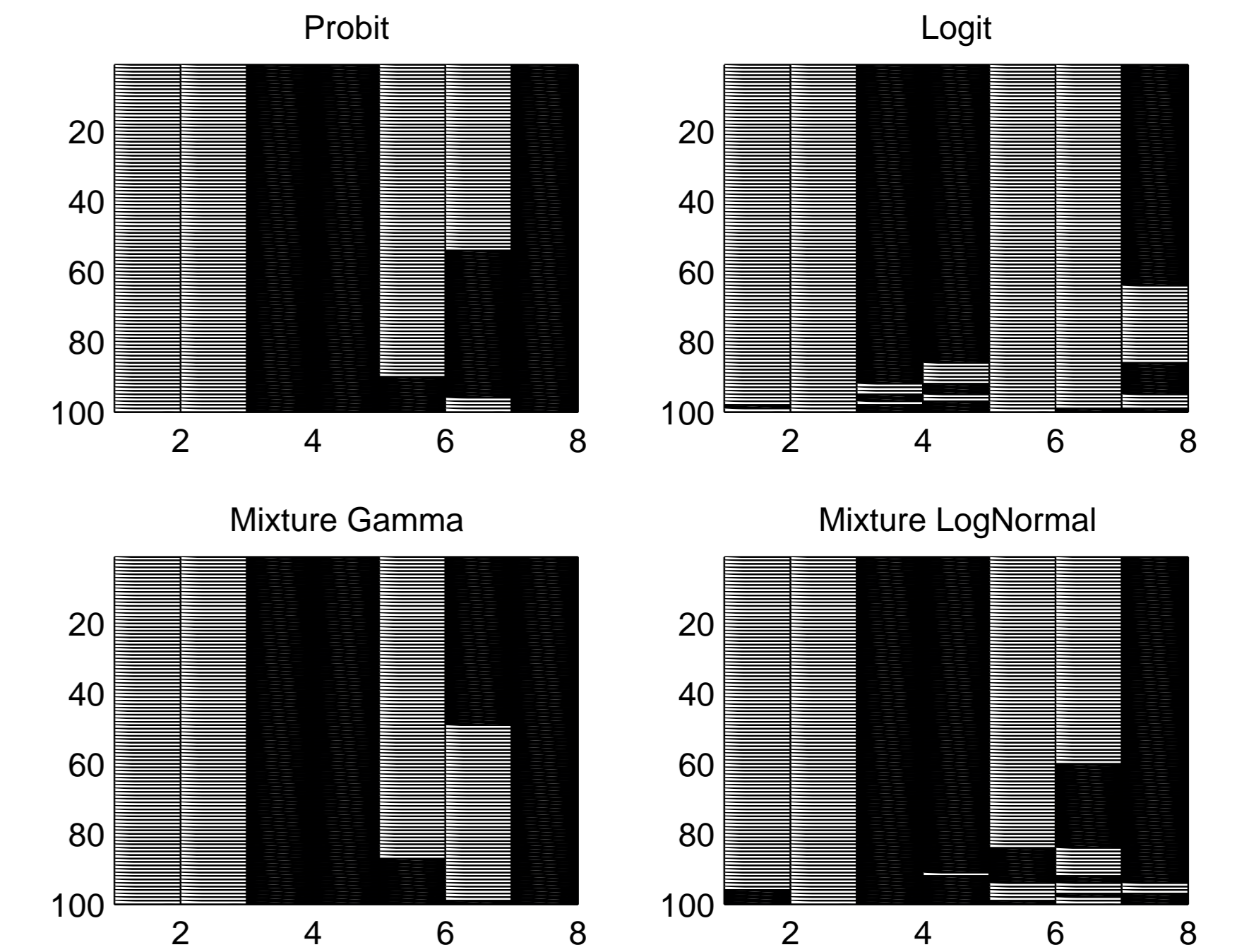


Figure 2: Posterior distribution  $\gamma$

## Conclusion

We presented a fully Bayesian approach to consider model uncertainty in link function and covariate set uncertainty when modelling binary data. The proposal is an extension to the regression model proposed by Albert and Chib (1993), a scale mixture of normals linear model on latent continuous data. We incorporated covariate set uncertainty by adding a hierarchy in our algorithm, where the update is done using the RJMCMC. We detailed the proposed model and the MCMC algorithm. The flexibility and power of the model was illustrated with an example. Our results suggest that the probability of including the covariate  $j$  depends, at least in some way, on the link function. This is interesting to consider, in practice, where the inclusion or exclusion of a covariate may dictate different policies, or lead to different conclusions.

## Acknowledgements

The first author is grateful to Consejo Nacional de Ciencia y Tecnología, México (CONACYT) for funding her research.

## References

- [1] ALBERT, J. H., AND CHIB, S. Bayesian analysis of binary data and polychotomous response data. *Journal of the American Statistical Association* 88, 422 (1993), 669–679.
- [2] FERNANDEZ, C., AND STEEL, M. Bayesian regression analysis with scale mixture of normals. *Econometric Theory* 16 (2000), 80–101.
- [3] GREEN, P. J. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 4 (1995), 711–732.
- [4] HOLMES, C. C., AND HELD, L. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* 1, 1 (2006), 145–168.