Regression models for misclassified binary data

Lizbeth Naranjo Albarrán

Jacinto Ramón Martín Jiménez

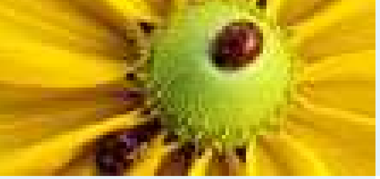
Carlos Javier Pérez Sánchez

Departamento de Matemáticas, Universidad de Extremadura

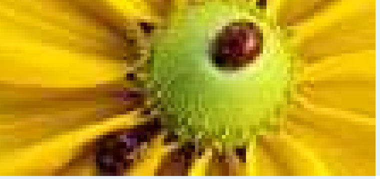
lizbeth@unex.es

Index

- Introduction
- The Bayesian model
- Illustrative example



Introduction



Generalized linear models

The generalized linear models proposed by Nelder and Wedderburn (1972) are specified by three components:

1. The *random component*: independent observations, $y_1, \ldots, y_n$, with distribution

$$f(y_i; \theta_i) = \exp\{(y_i\theta_i - b(\theta_i))/a_i(\phi) + c(y_i)\}$$

2. The *systematic component* or linear predictor:

$$\eta(\cdot) = \mathbf{x}\boldsymbol{\beta}$$

3. The *link function*: monotone and diferenciable function that describes the relation between the random and systematic components:

$$g(\mu_i) = \eta_i = \sum_{j=1}^{k} \beta_j x_{ij}, \qquad \text{where} \quad \mu_i = \mathrm{E}(y_i).$$

Generalized linear models for binary data

- The response variables have only two categories.
- Denote a binary response variable by Y and its two possible outcomes by 1 (“success”) and 0 (“failure”).
- We have $\mu_i = \mathrm{E}(y_i)$, where $0 < \mu_i < 1$.
- A link function should satisfy the condition that it maps the interval $(0, 1)$ over the whole real line.
  - Symmetric links
    - *logit*: $\eta = \log\{\mu/(1-\mu)\} = \Psi^{-1}(\mu)$, $\Psi(\cdot)$ logistic cdf
    - *probit*: $\eta = \Phi^{-1}(\mu)$, $\Phi(\cdot)$ normal cdf
    - *t-link*: $\eta = \Psi^{-1}(\mu)$, $\Psi(\cdot)$ t-Student cdf
  - Asymmetric links
    - *complementary log-log*: $\eta = \log\{-\log(1-\mu)\} = \Psi^{-1}(\mu)$, $\Psi(\cdot)$ Gumbel cdf
    - *skew probit*: $\eta = \Psi^{-1}(\mu)$, $\Psi(\cdot)$ skew normal cdf

Misclassified categorical data

- Classification appears naturally in many situations.
- When information is collected in the real word, the data are not usually free of error.
- This fact can happen due to several causes.
- Even a small proportion of misclassified data can produce an important impact on inferences.
- For example, in consumer surveys, consumers may:
  - not remember their previous behaviours accurately.
  - misunderstand survey questions.
  - intentionally misreport.
- Main consequence: important effects on the inferences.
- Noise or distortion must be statistically modelled.

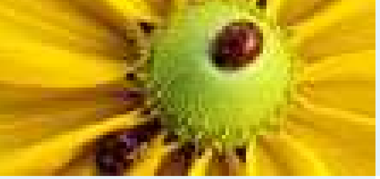
Proposed models

- GLM are used to describe the dependence of binary data on explanatory variables when the binary outcome is subject to misclassification.
- Statistical methodology: Bayesian.
- Precursors for logistic models:
  - Cowling et al. 2001.
  - Achcar et al. 2004.
  - Paulino et al. 2005.
- Proposed models: probit and t-link based regressions.
- Extension from Albert and Chib (1993) to address misclassification.

Probit and t-link models

- Main problem: computations.
- Solved by MCMC methods (Gibbs sampling).
- But previously, a data augmentation scheme is used.
- The model increases its dimensionality, but the generation process becomes easier.

The Bayesian model



Setting

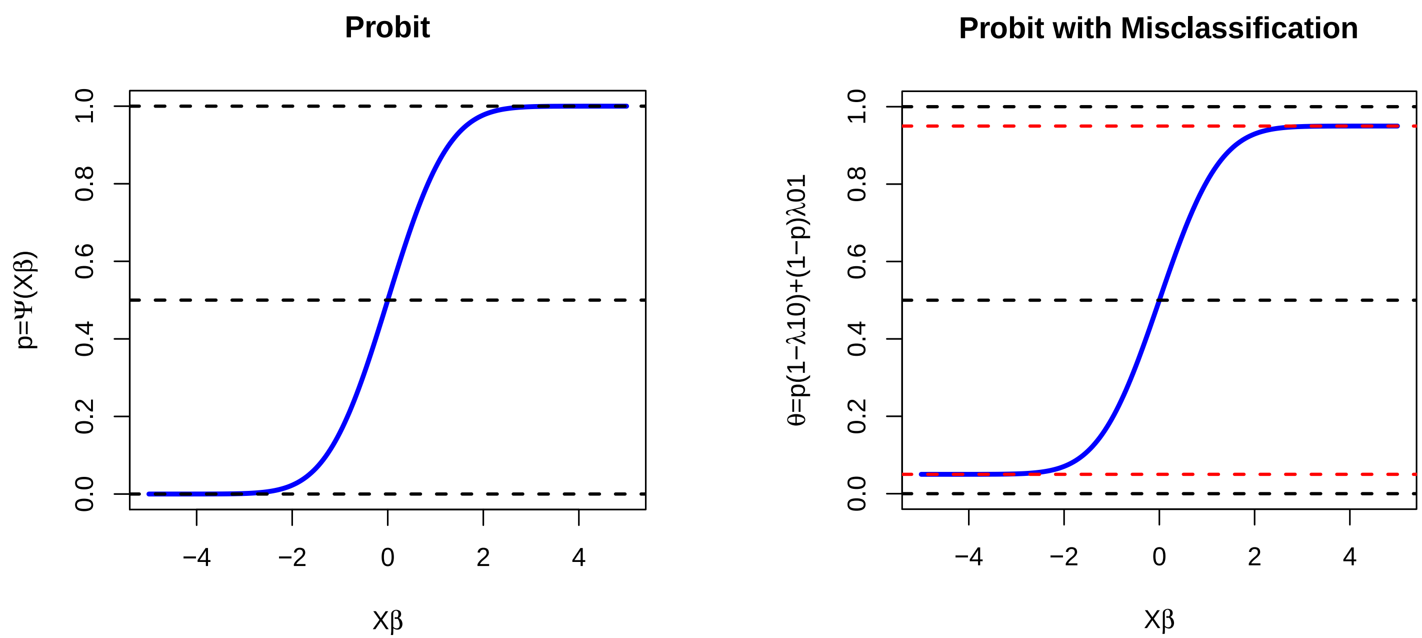
- $Y_1, \ldots, Y_n$ independent binary random variables.
- $Y_i \sim \text{Bernoulli}(p(Y_i = 1) = \theta_i)$
- θ_i is related to a covariate set $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^T$ through a binary regression with misclassification.
- Binary response model: $p_i = \Psi(x_i^T \boldsymbol{\beta})$
- $g(\cdot) = \Psi^{-1}$ is the link function.
- Ψ is a cumulative distribution function:
  - normal distribution (probit)
  - t-Student distribution (t-link).

Introducing noise parameters

- Misclassification is introduced in the model by:

$$\theta_i = p_i(1 - \lambda_{10}) + (1 - p_i)\lambda_{01}$$

  - p_i is the true positive probability for the observation i,
  - λ_{10} is the false negative probability,
  - λ_{01} is the false positive probability.



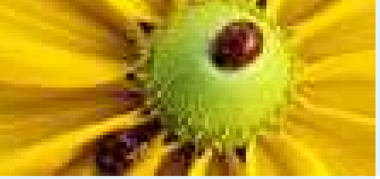
Introducing latent variables

- A data augmentation scheme is proposed.
- The first type of latent variables is c^i_{hk}, , $k = 0, 1$, where
  - $c^i_{11} = 1$ if i is a true positive,
  - $c^i_{10} = 1$ if i is a false negative,
  - $c^i_{01} = 1$ if i is a false positive,
  - $c^i_{00} = 1$ if i is a true negative.
- Latent vector and latent matrix:

$$\mathbf{c}^i = (c^i_{11}, c^i_{10}, c^i_{01}, c^i_{00})^T \qquad \mathbf{c} = (\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^n)^T$$

Introducing latent variables

- The second type of latent variables is introduced based on Albert and Chib (1993).
  - n independent latent variables $z_1, \ldots, z_n$ are considered, where z_i is distributed $\mathrm{N}(x_i^T \boldsymbol{\beta}, \gamma_i^{-1})$.
  - Define
$$\begin{cases} c_{11}^i + c_{10} = 1 & \text{if } z_i > 0 \\ c_{01}^i + c_{00} = 1 & \text{if } z_i \leq 0 \end{cases}$$
  - If the probit model is assumed, then $\gamma_i = p(\gamma_i) = 1$.
  - If the t-link model is assumed, then γ_i is distributes $\mathrm{Gamma}(\nu/2, 2/\nu)$, with pdf
$$p(\gamma_i) = c(\nu)\gamma_i^{\nu/2-1}\exp(-\nu\gamma_i/2),$$
where $c(\nu) = [\Gamma(\nu/2)(2/\nu)^{\nu/2}]^{-1}$.

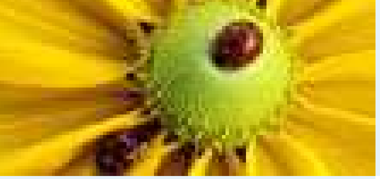


Likelihood function

- Likelihood function for the t-link model:

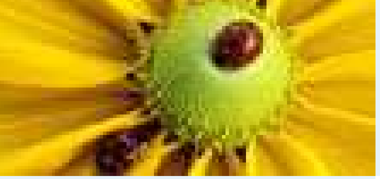
$$
\begin{aligned}
& L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \nu | \mathbf{D}) \\
\propto \quad & \prod_{i=1}^{n} \Big[\{p_i(1-\lambda_{10}) + (1-p_i)\lambda_{01}\}^{y_i} \\
\times \quad & \{p_i \lambda_{10} + (1-p_i)(1-\lambda_{01})\}^{1-y_i} \Big] \\
\propto \quad & \prod_{i=1}^{n} \Big[\int\int\int \Big\{ \phi(z_i; \mathbf{x}_i^T \boldsymbol{\beta}, \gamma_i^{-1}) \\
\times \quad & \left(I[z_i > 0] I[c_{11}^i + c_{10}^i = 1] + I[z_i \leq 0] I[c_{01}^i + c_{00}^i = 1] \right) \\
\times \quad & \left(I[y_i = 1] I[c_{11}^i + c_{01}^i = 1] + I[y_i = 0] I[c_{10}^i + c_{00}^i = 1] \right) \\
\times \quad & (1-\lambda_{10})^{c_{11}^i} \lambda_{10}^{c_{10}^i} \lambda_{01}^{c_{01}^i} (1-\lambda_{01})^{c_{00}^i} p(\gamma_i) \Big\} \mathrm{d}\gamma_i \mathrm{d}z_i \mathrm{d}\mathbf{c}^i \Big]
\end{aligned}
$$

- Probit model: particular case of the t-link based model.



Prior distributions

- The following step is defining the prior distributions.
  - Regression parameters β: as usual in error-free models (i.e. multivariate normal distribution), or based on the expert opinion (as in Bedrick et al. (1996)).
  - Noise parameters: with the natural choice for modelling the uncertainty about probabilities, i.e. Beta distributions.
  - Degrees of freedom: a bounded discrete distribution.
- These specifications allow to derive a Gibbs sampling algorithm to generate from the posterior distribution.
- All full conditional distributions can be efficiently generated by using standard algorithms.



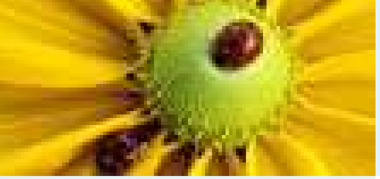
Prior distributions

- Bedrick et al. (1996) induce a prior probability distribution on $\boldsymbol{\beta}$ using a so called conditional means prior (CMP) on $\tilde{\mathbf{p}} = (\tilde{p}_1, \ldots, \tilde{p}_k)^T$, where in binomial regression $\tilde{p}_l = \mathrm{E}(\tilde{y}_l | \tilde{\mathbf{x}}_l)$ is the success probability for a potentially observable response $\tilde{y}_l$ at covariate vector $\tilde{\mathbf{x}}_l$.
  - Assuming k regression coefficients, prior probabilities $\tilde{p}_l$ are elicited in the predictor space, for selected locations $\tilde{\mathbf{x}}_l$.
  - With k linearly independent sets of covariate values, we obtain a 1-1 transformation between $\boldsymbol{\beta}$ and $\tilde{\mathbf{p}}$, namely $\boldsymbol{\beta} = \tilde{\mathbf{x}}^{-1}\Psi^{-1}(\tilde{\mathbf{p}})$, where $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1^T, \ldots, \tilde{\mathbf{x}}_k^T)^T$.
  - Uncertainty about $\tilde{p}_l$ is modelled with independent distributions $\mathrm{Be}(a_l, b_l)$. The hyperparameters a_l and b_l are determined from expert prior judgements.

Defining the prior distribution of $\boldsymbol{\beta}$ as $\mathrm{N}_k(\mathbf{b}_0, \mathbf{B}_0)$.

Given the data $\mathbf{D}$, the joint posterior distribution of the unobservables $\mathbf{c}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and ν is

$$
\begin{aligned}
& \pi(\mathbf{c}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu | \mathbf{D}) \propto \pi(\boldsymbol{\beta})\pi(\boldsymbol{\lambda})\pi(\nu) \\
\times \quad & \prod_{i=1}^{n} \Big[\{p_i(1-\lambda_{10})\}^{c_{11}^i} \{p_i \lambda_{10}\}^{c_{10}^i} \\
\times \quad & \{(1-p_i)\lambda_{01}\}^{c_{01}^i} \{(1-p_i)(1-\lambda_{01})\}^{c_{00}^i} \\
\times \quad & \big(I[y_i = 1] I[c_{11}^i + c_{01}^i = 1] + I[y_i = 0] I[c_{10}^i + c_{00}^i = 1] \big) \Big]
\end{aligned}
$$



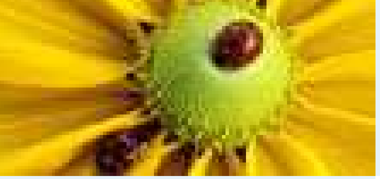
Posterior distributions (normal prior)

The full conditional distributions for $\mathbf{c}$ and $\boldsymbol{\lambda}$ are easy to obtain:

- $$\mathbf{c}^i|\boldsymbol{\beta},\boldsymbol{\lambda},\nu,\mathbf{D} \sim \text{Multinomial}\left(1, \pi_{c^i}(c^i_{11}, c^i_{10}, c^i_{01}, c^i_{00})\right),$$

$$\begin{aligned}
\pi_{c^i}(1,0,0,0) &= \frac{p_i(1-\lambda_{10})}{\theta_i} I[y_i = 1], \\
\pi_{c^i}(0,1,0,0) &= \frac{p_i\lambda_{10}}{(1-\theta_i)} I[y_i = 0], \\
\pi_{c^i}(0,0,1,0) &= \frac{(1-p_i)\lambda_{01}}{\theta_i} I[y_i = 1], \\
\pi_{c^i}(0,0,0,1) &= \frac{(1-p_i)(1-\lambda_{01})}{(1-\theta_i)} I[y_i = 0].
\end{aligned}$$

- $$\begin{aligned}
\lambda_{10}|\mathbf{c},\boldsymbol{\beta},\nu,\mathbf{D} &\sim \text{Be}\left(a_{10} + \sum_{i=1}^{n} c^i_{10},\ b_{10} + \sum_{i=1}^{n} c^i_{11}\right), \\
\lambda_{01}|\mathbf{c},\boldsymbol{\beta},\nu,\mathbf{D} &\sim \text{Be}\left(a_{01} + \sum_{i=1}^{n} c^i_{01},\ b_{01} + \sum_{i=1}^{n} c^i_{00}\right),
\end{aligned}$$



Posterior distributions (normal prior)

- The full conditional distributions $\pi(\boldsymbol{\beta}|\mathbf{c}, \boldsymbol{\lambda}, \nu, \mathbf{D})$ and $\pi(\nu|\mathbf{c}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{D})$ have not closed expression from which to generate easily.
- Generating from these distributions could be addressed by using a Metropolis-Hasting algorithm, however a Gibbs-within-Gibbs algorithm is more efficient and easier to implement by considering the introduction of latent variables in $\pi(\boldsymbol{\beta}, \nu|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$.
- The new distribution of interest is

$$
\begin{aligned}
& \pi(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \nu|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D}) \\
\propto \quad & \pi(\boldsymbol{\beta})\pi(\nu)\prod_{i=1}^{n}\Big\{\phi(z_i; \mathbf{x}_i^T\boldsymbol{\beta}, \gamma_i^{-1})p(\gamma_i) \\
\times \quad & \left(I[z_i > 0]I[c_{11}^i + c_{10}^i = 1] + I[z_i \leq 0]I[c_{01}^i + c_{00}^i = 1]\right)\Big\}.
\end{aligned}
$$

Posterior distributions (normal prior)

- The full conditional distributions of $z_1, \ldots, z_n$ are conditionally independent

$$z_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \nu, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D} \sim \begin{cases} \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta}, \gamma_i^{-1})I[z_i > 0] & \text{if } c_{11}^i + c_{10}^i = 1 \\ \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta}, \gamma_i^{-1})I[z_i \leq 0] & \text{if } c_{01}^i + c_{00}^i = 1 \end{cases}.$$

- $\boldsymbol{\beta}$ is obtained by

$$\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\gamma}, \nu, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D} \sim \mathrm{N}_k\left(\mathbf{b}_k, \mathbf{B}_k\right),$$

where

$$\mathbf{b}_k = \mathbf{B}_k(\mathbf{x}^T\mathrm{W}\mathbf{z} + \mathbf{B}_0^{-1}\mathbf{b}_0), \quad \mathbf{B}_k = (\mathbf{x}^T\mathrm{W}\mathbf{x} + \mathbf{B}_0^{-1})^{-1},$$

and $\mathrm{W} = \mathrm{diag}(\gamma_i)$.

For the t-link model:

- The full conditional distributions of $\gamma_1, \ldots, \gamma_n$ are conditionally independent with

$$\gamma_i | \mathbf{z}, \boldsymbol{\beta}, \nu, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D} \sim \mathrm{Ga}\left(\frac{\nu+1}{2}, \frac{2}{\nu + (z_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}\right).$$

- $\nu | \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D}$ is distributed according to a pmf proportional to

$$\pi(\nu) \prod_{i=1}^{n} \left(c(\nu) \gamma_i^{\nu/2-1} e^{-\nu\gamma_i/2}\right).$$

Posterior distributions (eliciting prior)

Building a prior distribution for $\boldsymbol{\beta}$ based on the expert prior elicitation was proposed by Bedrick et al. (1996).

- The independence CMP

$$\pi(\tilde{\mathbf{p}}) \propto \prod_{l=1}^{k} \tilde{p}_l^{a_l-1}(1-\tilde{p}_l)^{b_l-1},$$

induces a prior on $\boldsymbol{\beta}$ given by

$$\pi(\boldsymbol{\beta}) \propto \prod_{l=1}^{k} \Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})^{a_l-1}[1-\Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})]^{b_l-1}\psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta}),$$

where $\Psi = \Phi$ for the probit model and $\Psi = T_\nu$ for the t-link model.

- The posterior distribution is

$$\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \nu|\mathbf{D}) \propto \pi(\boldsymbol{\beta})\pi(\boldsymbol{\lambda})\pi(\nu)L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \nu).$$

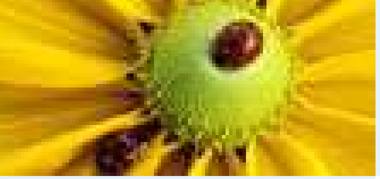
Posterior distributions (eliciting prior)

- To sample from $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and ν, the latent variables $\mathbf{c}$ are introduced. The full conditional distributions for $\mathbf{c}$ and $\boldsymbol{\lambda}$ are the same as in case of normal prior.

- The full conditional distributions to $\boldsymbol{\beta}$ and ν are given by

$$\begin{aligned}\pi(\boldsymbol{\beta}|\mathbf{c},\boldsymbol{\lambda},\nu,\mathbf{D}) &\propto \prod_{l=1}^{k} \Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})^{a_l-1}[1-\Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})]^{b_l-1}\psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta}) \\ &\times \prod_{i=1}^{n} \Psi(\mathbf{x}_i^T\boldsymbol{\beta})^{c_{11}^i+c_{10}^i}[1-\Psi(\mathbf{x}_i^T\boldsymbol{\beta})]^{c_{01}^i+c_{00}^i},\end{aligned}$$

$$\begin{aligned}\pi(\nu|\mathbf{c},\boldsymbol{\lambda},\boldsymbol{\beta},\mathbf{D}) &\propto \pi(\nu)\prod_{l=1}^{k} \Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})^{a_l-1}[1-\Psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta})]^{b_l-1}\psi(\tilde{\mathbf{x}}_l^T\boldsymbol{\beta}) \\ &\times \prod_{i=1}^{n} \Psi(\mathbf{x}_i^T\boldsymbol{\beta})^{c_{11}^i+c_{10}^i}[1-\Psi(\mathbf{x}_i^T\boldsymbol{\beta})]^{c_{01}^i+c_{00}^i}.\end{aligned}$$



Illustrative example

Context

Index
Introduction
The Bayesian model
Illustrative example
- Context
- Misclassified cases
- Prior
- Predictions and DIC
- Conclusions

References

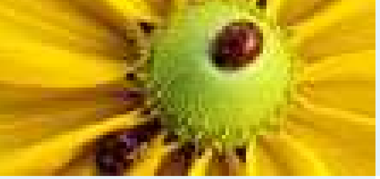
- A covariate set is generated by $x_{i1} \sim \mathrm{N}(2, 0.09)$, and $x_{i2} \sim \mathrm{N}(3, 0.09)$, $i = 1, \ldots, 100$.
- The probabilities are obtained for both error-free models by:

$$\eta_i = \Psi^{-1}(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2},$$

  where $\boldsymbol{\beta} = (2, -4, 2)^T$.
- For each model, the true binary dependent variable y^{true} is obtained by

$$\begin{cases} y_i^{true} = 0 & \text{if} \quad p_i \leq 0.5 \\ y_i^{true} = 1 & \text{if} \quad p_i > 0.5 \end{cases}$$

- Both probit and $t(7)$-Student models display the same outcomes because of the symmetric links and the discretization.

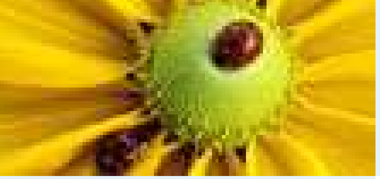


Context

- Intentionally misclassify some outcomes according to the following quantities:
  - 7 outcomes with $y^{true} = 0$ becoming $y = 1$,
  - 5 outcomes with $y^{true} = 1$ becoming $y = 0$.
- The new response variable y remains equal to y^{true} for the non-misclassified outcomes.
- Then, the known proportion of misclassification for the new response is given by:

$$\lambda_{01} = \mathrm{p}(\text{false positive}) = 1 - \text{Specificity} = \frac{7}{45} = 0.1555$$

$$\lambda_{10} = \mathrm{p}(\text{false negative}) = 1 - \text{Sensitivity} = \frac{5}{55} = 0.0909$$



Misclassified cases

- Three misclassification cases are considered for both models, according to the previous quantities.

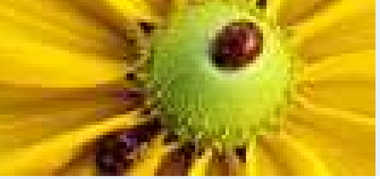
  1. Only outcomes close to the "border" are misclassified, i.e. outcomes for which $p_i \approx 0.5$.

  2. Outcomes being far from the "border", i.e. $p_i \approx 0$ or $p_i \approx 1$.

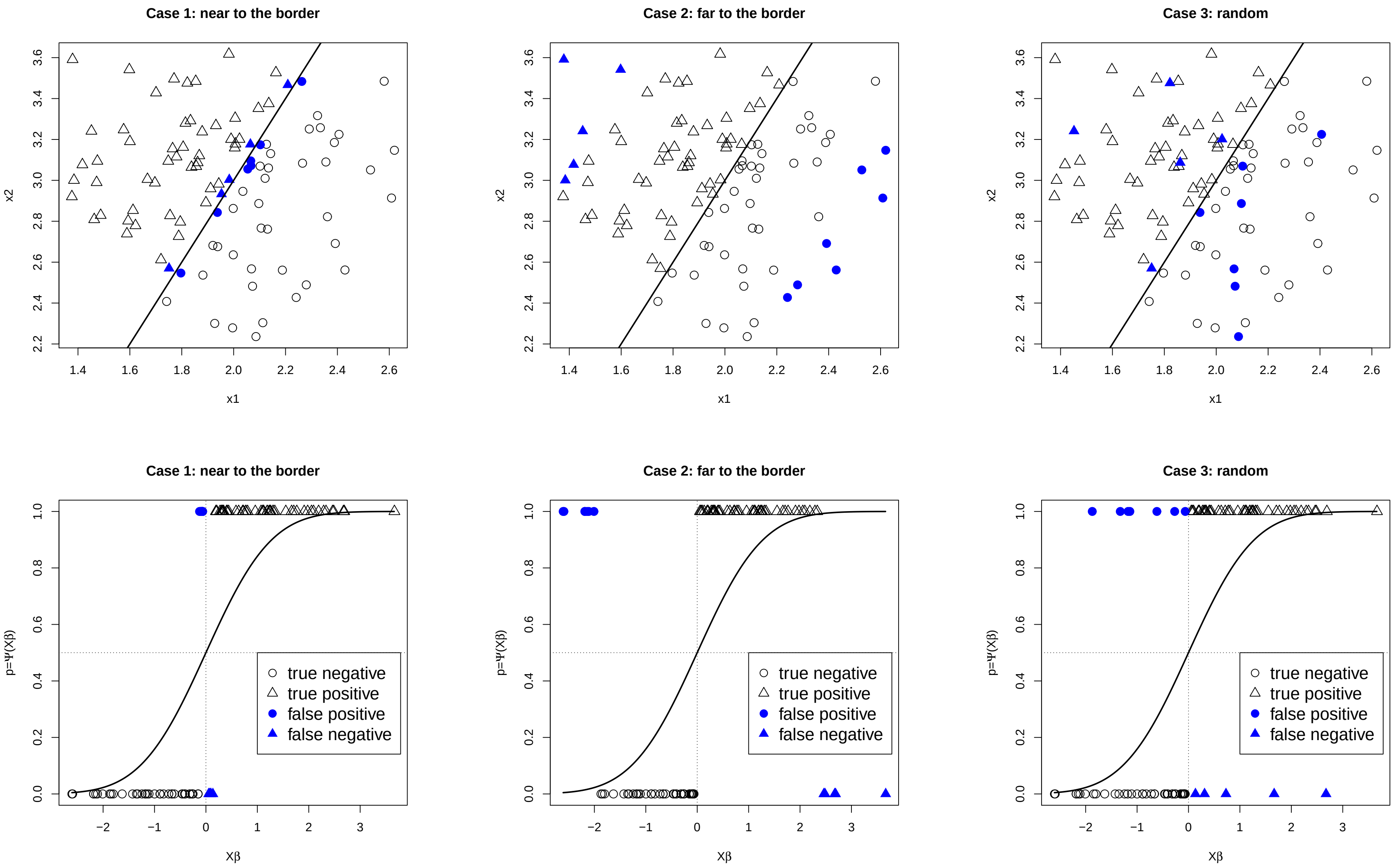
  3. A random misclassification is considered.

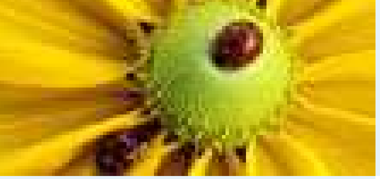
- The main objective is to compare the predictive performance of the proposed models to the standard ones.

- This simulated scenario allows to compare the predictive outcomes with the real ones and, therefore, to know what model performs better.



Misclassified cases





Prior

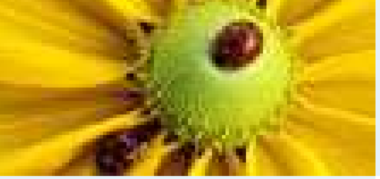
- Three prior specifications are considered for the regression parameters $\boldsymbol{\beta}$:

  1. A informative prior, $\boldsymbol{\beta} \sim \mathrm{N}_3(\mathbf{b}_0^T, \mathbf{B}_0)$, where $\mathbf{b}_0^T = (2, -4, 2)$ and $\mathbf{B}_0 = \mathrm{diag}(10, 10, 10)$.

  2. A weakly informative prior, $\boldsymbol{\beta} \sim \mathrm{N}_3(\mathbf{b}_0^T, \mathbf{B}_0)$, where $\mathbf{b}_0^T = (0, 0, 0)$ and $\mathbf{B}_0 = \mathrm{diag}(100, 100, 100)$.

  3. A prior based on the proposal of Bedrick et al. (1996), that have been adapted to address misclassifications:

Configurations	Hyperparameters	
$\tilde{\mathbf{x}}_1^T = (1.7, 3.3)$	$a_{11} = 45$	$a_{21} = 5$
$\tilde{\mathbf{x}}_2^T = (2.0, 3.0)$	$a_{12} = 25$	$a_{22} = 25$
$\tilde{\mathbf{x}}_3^T = (2.3, 3.3)$	$a_{13} = 5$	$a_{23} = 15$
λ_{10}	$a_{10} = 5$	$b_{10} = 50$
λ_{01}	$a_{01} = 7$	$b_{01} = 38$



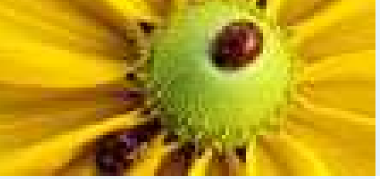
Prior

- For the noise parameters, we introduce initial information according to the misclassification proportions that affect the data, i.e.:

$$\lambda_{10} \sim \mathrm{Be}(5, 50) \qquad \text{and} \qquad \lambda_{01} \sim \mathrm{Be}(7, 38)$$

- Two cases are considered for the distribution of the degrees of freedom:

1. $\nu = 7$,
2. ν being a r.v. with discrete finite distribution. Its support is $\{4, \ldots, 10\}$ and the probabilities are $(0.05, 0.10, 0.20, 0.30, 0.20, 0.10, 0.05)$.



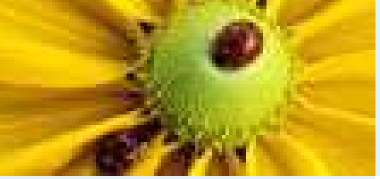
Predictions and DIC

Case 1: Misclassification near to the border

Predictions	$y^{true}=0$ $y^{pred}=0$	$y^{true}=0$ $y^{pred}=1$	$y^{true}=1$ $y^{pred}=0$	$y^{true}=1$ $y^{pred}=1$
Probit	42/43/45	3/2/0	0/0/1	55/55/54
Probit Mis.	42/43/45	3/2/0	0/0/0	55/55/55
$t(7)$	42/43/45	3/2/0	0/0/0	55/55/55
$t(7)$ Mis.	42/43/45	3/2/0	0/0/0	55/55/55
$t(r.v.)$	42/43/45	3/2/0	0/0/0	55/55/55
$t(r.v.)$ Mis.	42/43/45	3/2/0	0/0/0	55/55/55

Informative prior/ Weakly informative prior/ Elicited prior information

DIC	Informative prior	Weakly informative prior	Elicited prior information
Probit	37.486	35.012	55.015
Probit Mis.	48.144	45.904	65.407
$t(7)$	38.627	35.740	51.060
$t(7)$ Mis.	49.556	46.306	61.627
$t(r.v.)$	38.652	35.756	50.982
$t(r.v.)$ Mis.	49.631	46.418	56.966



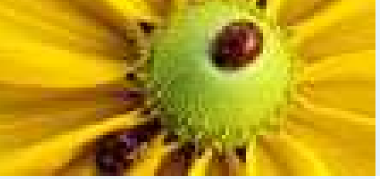
Predictions and DIC

Case 2: Misclassification far to the border

Predictions	$y^{true}=0$ $y^{pred}=0$	$y^{true}=0$ $y^{pred}=1$	$y^{true}=1$ $y^{pred}=0$	$y^{true}=1$ $y^{pred}=1$
Probit	31/30/36	14/15/9	0/0/0	55/55/55
Probit Mis.	44/45/45	1/0/0	0/0/0	55/55/55
$t(7)$	32/32/37	13/13/8	0/0/0	55/55/55
$t(7)$ Mis.	44/45/44	1/0/1	0/0/0	55/55/55
$t(r.v.)$	32/32/37	13/13/8	0/0/0	55/55/55
$t(r.v.)$ Mis.	44/45/45	1/0/0	0/0/0	55/55/55

Informative prior/ Weakly informative prior/ Elicited prior information

DIC	Informative prior	Weakly informative prior	Elicited prior information
Probit	131.836	132.058	133.301
Probit Mis.	97.179	87.822	116.479
$t(7)$	131.656	131.880	132.675
$t(7)$ Mis.	99.813	89.219	113.194
$t(r.v.)$	131.728	131.740	132.686
$t(r.v.)$ Mis.	99.954	89.531	114.228



Predictions and DIC

Case 3: Misclassification random

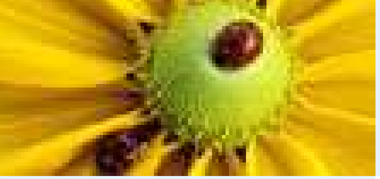
Predictions	$y^{true}=0$ $y^{pred}=0$	$y^{true}=0$ $y^{pred}=1$	$y^{true}=1$ $y^{pred}=0$	$y^{true}=1$ $y^{pred}=1$
Probit	41/41/42	4/4/3	0/0/1	55/55/54
Probit Mis.	43/44/44	2/1/1	0/0/0	55/55/55
$t(7)$	39/40/42	6/5/3	0/0/1	55/55/54
$t(7)$ Mis.	43/44/43	2/1/2	0/0/0	55/55/55
$t(r.v.)$	39/40/42	6/5/3	0/0/1	55/55/54
$t(r.v.)$ Mis.	43/44/44	2/1/1	0/0/1	55/55/54

Informative prior/ Weakly informative prior/ Elicited prior information

DIC	Informative prior	Weakly informative prior	Elicited prior information
Probit	98.848	99.306	96.792
Probit Mis.	86.408	82.027	95.453
$t(7)$	97.172	97.897	95.240
$t(7)$ Mis.	88.732	83.830	93.898
$t(r.v.)$	96.956	97.723	95.189
$t(r.v.)$ Mis.	88.672	84.027	94.462

Conclusions

1. Case 1: Misclassified outcomes are close to the border.
   - The models addressing misclassification perform similar to the standard models.
   - The proposed models are not able to identify the misclassified outcomes, giving similar predictions as in the standard models.
2. Case 2: Misclassified outcomes are far from the border.
   - The proposed models are able to identify and make right predictions, by relocating misclassified outcomes to their right values.
3. Case 3: Misclassified outcomes randomness.
   - The proposed models performs better than the standard one, but not as well as in the second case.



Conclusions

In general, the proposed methods:

- They are no worse than the standard methods.
- They can increase substantially the number of correct predictions with respect to the standard methods.
- They can produce better fits than the standard models.
- The use of latent variables as proposed here enables us to avoid computational difficulties.
- Although the augmented models increase the dimensionality, the generation processes become easier.

Thank you

References

[1] Achcar, J. A., Martínez, E. Z., and Louzada-Neto, F. (2004). *Proceedings of the Computational Statistics Conference*, chapter Binary data in the presence of misclassifications, pages 581–588. Physica-Verlag.

[2] Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

[3] Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association* **91**, 1450–1460.

[4] Cowling, D., Johnson, W. O., and Gardner, I. A. (2001). Bayesian modelling of risk when binary outcomes are subject to error. Technical report, Department of Statistics, University of California, Davis.

[5] Nelder, J., and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Socciety* **A135**, 370–384.

[6] Paulino, C. D., Silva, G., and Achcar, J. A. (2005). Bayesian analysis of correlated misclassified binary data. *Computational Statistics and Data Analysis* **49**, 1120–1131.