

SEMPARAMETRIC PROBIT REGRESSION MODEL FOR BINARY DATA

Brian M. Brost

07 August 2015

Description

A semiparametric, probit regression model for binary data.

Implementation

The file `probit.semipar.sim.R` simulates data according to the model statement presented below, and `probit.semipar.mcmc.R` contains the MCMC algorithm for model fitting.

Model statement

Let y_t , for $t = 1, \dots, T$, be observed data that take on the values $\{0, 1\}$. Also let \mathbf{X} be a design matrix containing covariates for which inference is desired, and \mathbf{Z} be a design matrix containing some basis expansion. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are the corresponding 'fixed' and 'random' effects, respectively. Note that $\mathbf{Z}\boldsymbol{\alpha}$ models non-linear patterns or dependence non-parametrically.

$$\begin{aligned} y_t &\sim \begin{cases} 0, & u_t \leq 0 \\ 1, & u_t > 0 \end{cases} \\ u_t &\sim \mathcal{N}(\mathbf{x}_t' \boldsymbol{\beta} + \mathbf{z}_t' \boldsymbol{\alpha}, 1) \\ \boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, \sigma_\beta^2 \mathbf{I}) \\ \boldsymbol{\alpha} &\sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}) \end{aligned}$$

Models of this type are typically fit using a large number of basis vectors, more than necessary to approximate non-linear trends or dependence. Regularization (e.g., a ridge penalty) is subsequently conducted to shrink the coefficients $\boldsymbol{\alpha}$ toward 0 where appropriate. Therefore, the parameter σ_α^2 must be selected using cross-validation or some model selection criterion.

Full conditional distributions

Auxiliary variable for binary observation model (u_t):

$$\begin{aligned} [u_t | \cdot] &\propto [y_t | u_t] [u_t] \\ &\propto (1_{\{y_t=0\}} 1_{\{u_t \leq 0\}} + 1_{\{y_t=1\}} 1_{\{u_t > 0\}}) \times \mathcal{N}(u_t | \mathbf{x}_t' \boldsymbol{\beta} + \mathbf{z}_t' \boldsymbol{\alpha}, 1) \\ &= \begin{cases} \mathcal{TN}(\mathbf{x}_t' \boldsymbol{\beta} + \mathbf{z}_t' \boldsymbol{\alpha}, 1)_{-\infty}^0, & y_t = 0 \\ \mathcal{TN}(\mathbf{x}_t' \boldsymbol{\beta} + \mathbf{z}_t' \boldsymbol{\alpha}, 1)_{0}^{\infty}, & y_t = 1 \end{cases} \end{aligned}$$

Fixed effects (β):

$$\begin{aligned}
[\beta|\cdot] &\propto [\mathbf{u}|\beta, \alpha, \sigma^2][\beta] \\
&\propto \mathcal{N}(\mathbf{u}|\mathbf{X}\beta + \mathbf{Z}\alpha, \mathbf{1})\mathcal{N}(\beta|\mathbf{0}, \sigma_\beta^2\mathbf{I}) \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{u} - (\mathbf{X}\beta + \mathbf{Z}\alpha))'(\mathbf{u} - (\mathbf{X}\beta + \mathbf{Z}\alpha))\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta - \mathbf{0})'(\sigma_\beta^2\mathbf{I})^{-1}(\beta - \mathbf{0})\right\} \\
&\propto \exp\left\{-\frac{1}{2}((\mathbf{u} - \mathbf{Z}\alpha) - \mathbf{X}\beta)'((\mathbf{u} - \mathbf{Z}\alpha) - \mathbf{X}\beta)\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta - \mathbf{0})'(\sigma_\beta^2\mathbf{I})^{-1}(\beta - \mathbf{0})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(-2(\mathbf{u} - \mathbf{Z}\alpha)'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta)\right\} \times \\
&\quad \exp\left\{-\frac{1}{2}(\beta'(\sigma_\beta^2\mathbf{I})^{-1}\beta)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2((\mathbf{u} - \mathbf{Z}\alpha)'\mathbf{X})\beta + \beta'(\mathbf{X}'\mathbf{X} + (\sigma_\beta^2\mathbf{I})^{-1})\beta\right)\right\} \times \\
&= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1})
\end{aligned}$$

where $\mathbf{A} = \mathbf{X}'\mathbf{X} + (\sigma_\beta^2\mathbf{I})^{-1}$ and $\mathbf{b}' = (\mathbf{u} - \mathbf{Z}\alpha)'\mathbf{X}$.

Random effects (α):

$$\begin{aligned}
[\alpha|\cdot] &\propto [\mathbf{u}|\beta, \alpha, \sigma^2][\alpha] \\
&\propto \mathcal{N}(\mathbf{u}|\mathbf{X}\beta + \mathbf{Z}\alpha, \mathbf{1})\mathcal{N}(\alpha|\mathbf{0}, \sigma_\alpha^2\mathbf{I}) \\
&= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1})
\end{aligned}$$

where $\mathbf{A} = \mathbf{Z}'\mathbf{Z} + (\sigma_\alpha^2\mathbf{I})^{-1}$ and $\mathbf{b}' = (\mathbf{u} - \mathbf{X}\beta)'\mathbf{Z}$.