On implementing new link functions for the VGAM R package

C. Somchit[1,2] and T. W. Yee[1]

[1]University of Auckland, New Zealand. [2]University of Phayao, Thailand.

Introduction

In the binary response model the probability, π_i of the success of an event, $Y_i = 1$, conditional on the covariate vector $(x_1, \ldots, x_p)$ is expressed as,

$$g(\pi_i) \;=\; \eta_i \;=\; \sum_{k=1}^{p} x_{ik}\,\beta_k\,; \quad i = 1, \ldots, n.$$

The function g links the linear predictors to the probability and determines the shape of the quantal response. McCullagh and Nelder [3] represent four possible link functions for binary response model:

$$\begin{aligned} \text{logit} &: \log\left(\pi/\left(1-\pi\right)\right), \\ \text{probit} &: \Phi^{-1}(\pi), \\ \text{cloglog} &: \log\left(-\log\left(1-\pi\right)\right), \\ \text{loglog} &: -\log\left(-\log\left(\pi\right)\right). \end{aligned}$$

Koenker[2] implements two parametric families of link functions for binary response applications:

Gosset link : the Student t link with a degree of freedom parameter, ν; $(\nu > 0)$.

$$\text{Pregibon link (Pregibon[4])} : g(\mu, a, b) \;=\; \frac{\mu^{a-b}-1}{a-b} - \frac{(1-\mu)^{a+b}-1}{a+b}. \qquad (1)$$

The VGAM package provides a number of link functions (Yee[5]) and in this article we implement the Student t link function to the VGAM package. Student t link is an alternative link function for parameters that lie in the unit interval. The link is described with a degree of freedom parameter, ν. When $\nu = 1$ the link is equivalent to cauchit link, while $\nu \to \infty$ this corresponds to probit link.

The derivations

The link function is denoted by

$$\eta \;=\; g(\theta) \;=\; \beta^T x.$$

The Student t link function uses the distribution function of the Student t random variables as the choice of the inverse of the link function. Therefore $\eta = F_\nu^{-1}(\theta)$ and $\theta = F_\nu(\eta)$, where $\theta \in (0,1)$ is the vector of parameters to be estimated. The first two derivatives of θ with respect to η are

$$\begin{aligned} \frac{\partial\theta}{\partial\eta} &= F_\nu'\left(F_\nu^{-1}(\theta)\right) \;=\; f_\nu\left(F_\nu^{-1}(\theta)\right), \\ \frac{\partial^2\theta}{\partial\eta^2} &= \frac{\partial}{\partial\eta} F_\nu'\left(F_\nu^{-1}(\theta)\right) \;=\; \frac{\partial}{\partial\eta} f_\nu\left(F_\nu^{-1}(\theta)\right) \;=\; -\left(F_\nu^{-1}(\theta)\right)\cdot\left(f_\nu\left(F_\nu^{-1}(\theta)\right)\right). \end{aligned}$$

The Student t link function

The structure of `stlink()` consists of components for computing the link function and its first and second derivative, the inverse link and its first and second derivative, and the character description of the link. The link function has required the degree of freedom as a specified parameter.

The Student t link function

```
stlink(theta, earg = list(df = Inf), inverse = FALSE, deriv = 0,
       short = TRUE, tag = FALSE)
```

The accuracy

```
> all.equal(stlink(p), probit(p))
[1] TRUE
> all.equal(stlink(p, earg=list(df=1)), cauchit(p))
[1] TRUE
> all.equal(stlink(stlink(p), inverse = TRUE), p)
[1] TRUE
```

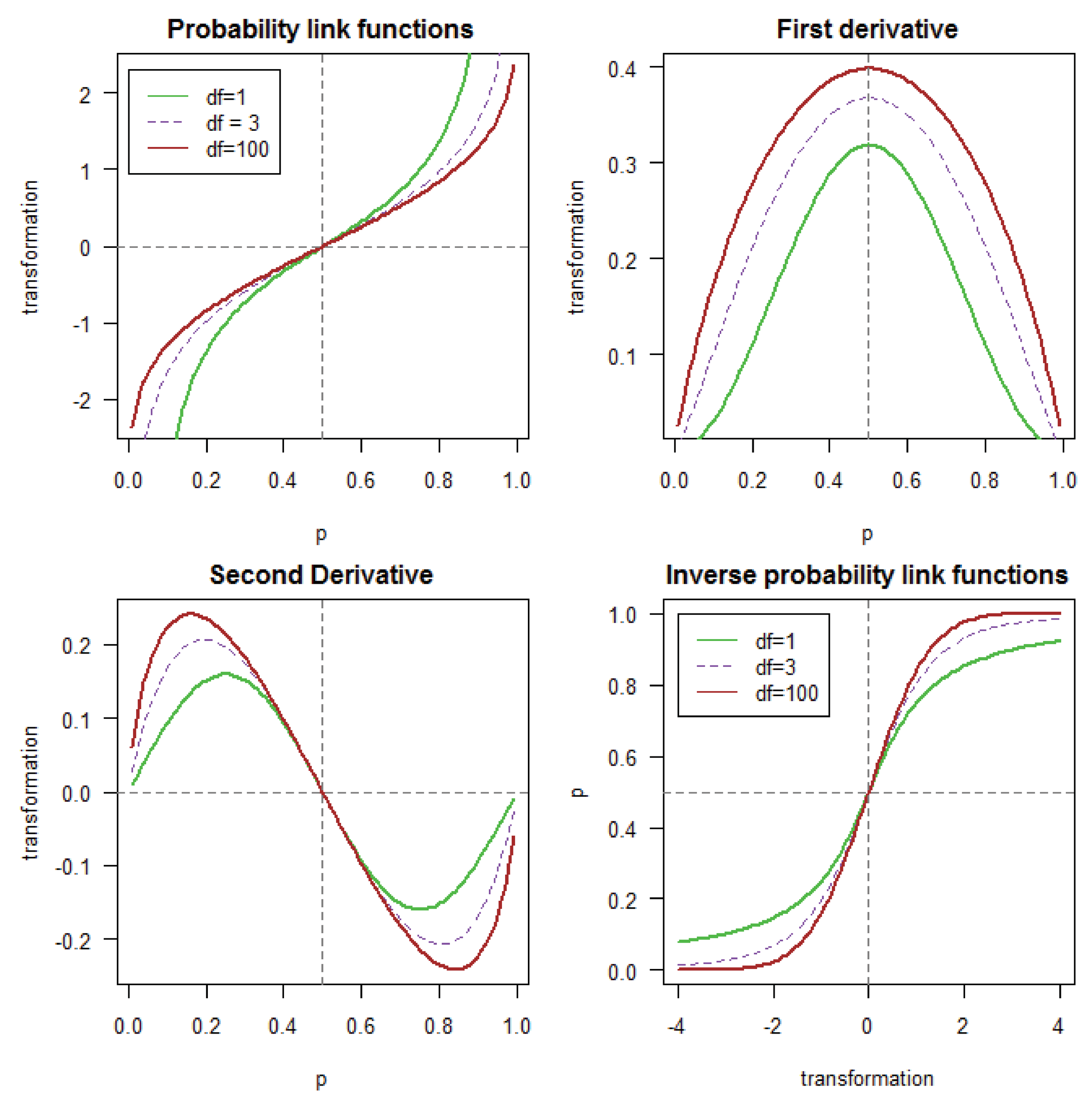

FIGURE 1: Student t link values (top left) and first two derivative (top right and bottom left respectively), and inverse values (bottom right) for various values of `df`.

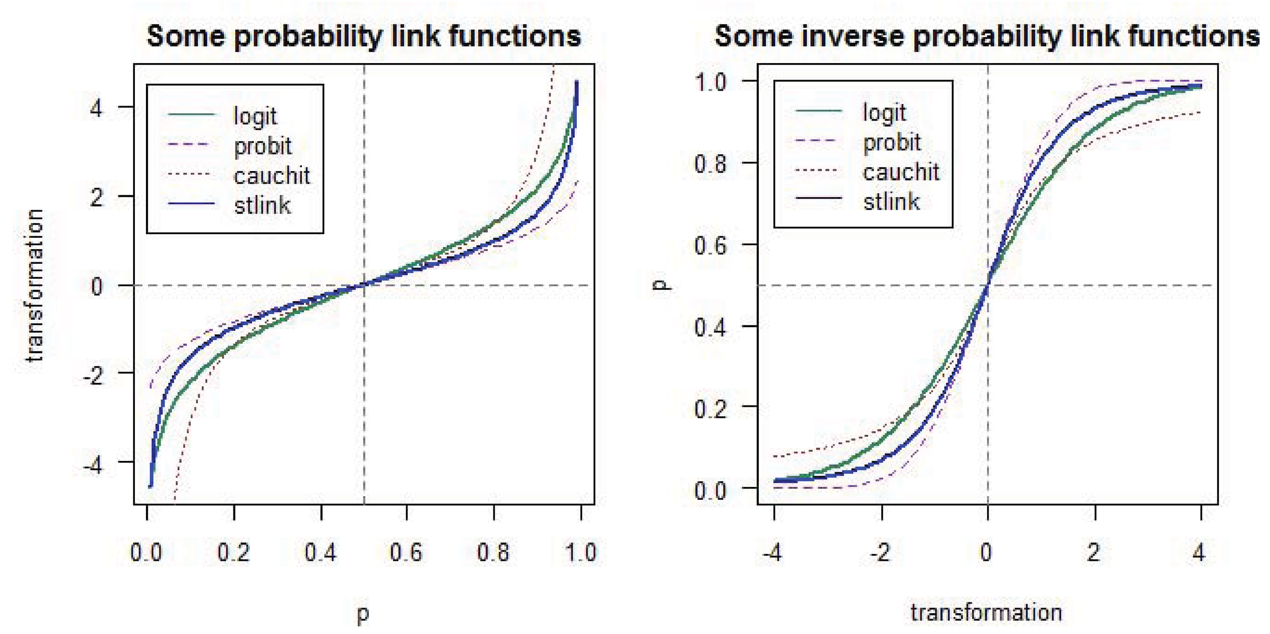

FIGURE 2: Values of four links, `stlink` with `df = 3` (left) and inverse values (right).

AIDS and AZT Data (Agresti[1]) The coefficients, log-likelihood, AIC , and the deviance obtained from `glm()` and `vglm()` with `stlink()` comparing to the corresponding link functions.

```
> all.equal(coef(Gosset), coef(stlink))
[1] TRUE
```

Link functions	Log-likelihood	AIC	Deviance
`logit()`	-9.4299	24.8598	1.3835
`Gosset(df = Inf)`	-9.4383	24.8765	1.4003
`probit()`	-9.4383	24.8765	1.4003
`stlink(df = Inf)`	-9.4383	24.8765	1.4003
`Gosset(df = 1)`	-9.3339	24.6677	1.1914
`cauchit()`	-9.3339	24.6677	1.1914
`stlink(df = 1)`	-9.3339	24.6677	1.1914

Table 1: Log-likelihood, AIC, and deviance statistic of models using `stlink()` and other corresponding link functions for the AIDS and AZT data sets.

Future work

The next step is to investigate and develop a wider class of parametric link functions for binary and ordinal responses.

Link functions for binary responses

The logit and probit are widely used in modeling for binary response data. However they do not always provide the best fit available for a given data set. In the future frame work we will investigate and implement more flexible link functions such as

the Pregibon (1), the power logit : $\text{logit}(\theta^a)$, and nloglog : $-\log(-\log(\theta))$.

Link functions for ordinal responses

Yee[6] develops three new link functions for ordinal responses obtained from the Poisson or negative binomial distribution : a Poisson-ordinal link function(POLF), a negative binomial-ordinal link function (NBOLF), and a 2-parameter gamma(GOLF).

$$\begin{aligned} \text{POLF} &: 2\log\left(Q_l/2 + \sqrt{K_l + \tfrac{7}{8}}\right), \quad K_l > 0, \\ &: \log\left(-\log\left(1-\gamma_1^c\right)\right), \quad K_1 = 0, \\ \text{NBOLF} &: 2\log\left(\sqrt{k}\,\sinh\left(\frac{Q_l}{2\sqrt{k}} + \sinh^{-1}\sqrt{\frac{K_l}{k}}\right)\right), \quad K_l > 0, \\ &: \log k + \log\left(\left(1-\gamma_1^c\right)^{-1/k} - 1\right), \quad K_1 = 0, \\ \text{GOLF} &: -3\log\left(1 - Q_l/\left(3\sqrt{\lambda}\right)\right) + \log K_l. \end{aligned}$$

In this area we will improve the methodology of Yee[6] to develop new link functions for Poisson and negative binomial distribution that a random variable Y has been observed as an ordinal response.

References

[1] A. Agresti. *Categorical Data Analysis*. New York: John Wiley & Sons, 2nd edition edition, 2002.

[2] R. Koenker. Parametric links for binary response. *R News*, 6(4):32–34, October 2006.

[3] P. McCullagh and J.A. Nelder. *Generalized linear Models*. London: Chapman & Hall, 2nd edition edition, 1989.

[4] D. Pregibon. Goodness of link tests for generalized linear models. *Applied Statistics*, 29(1):15–23, 1980.

[5] T. W. Yee. Writing VGAM family functions. 2008.

[6] T. W. Yee. Ordinal ordination with normalizing link functions for count data. *In preparation*, 2011.