Power logit regression for modeling bounded continuous data

Francisco F. Queiroz, Silvia L. P. Ferrari
University of São Paulo

Abbreviated abstract: We introduce a new class of regression models for bounded continuous data, such as continuous proportions, scores and rates. The models, named the power logit regression models, assume that the response variable follows a distribution in a wide, flexible class of distributions with three parameters, namely the median, a dispersion parameter and a skewness parameter. We offer a comprehensive set of tools for likelihood inference and diagnostic analysis, and introduce the new R package PLreg. Applications with real and simulated data show the merits of the proposed models, the statistical tools, and the computational package.

Related publications:

– Ferrari and Cribari-Neto, Journal of Applied Statistics, 31, 799-815 (2004)
– Lemonte and Bazán, Biometrical Journal, 58, 727–746 (2016)

IME-USP

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Introduction and previous works

- **Context**: predict or explain the behavior of a continuous proportion from a set of other variables.

Natural approach: regression model where the dependent variable has a probability distribution on (0, 1).

- Beta regression models (Ferrari and Cribari-Neto, 2004).
  - simple;
  - direct parameter interpretation;
  - software available.
  - Drawbacks: likelihood inference is usually influenced by atypical observations; not flexible enough to accommodate some patterns of data.

- GJS regression models.
  - Lemonte and Bazan (2016) define the GJS distributions by assuming that

$$\frac{1}{\sigma}[\text{logit}(Y) - \text{logit}(\mu)] \sim \text{S}(0, 1; r),$$

$0 < \mu < 1$, $\sigma > 0$, and $\text{S}(0, 1; r)$ denotes a standardized symmetric distribution (e.g. normal, Student-t, power exponential).

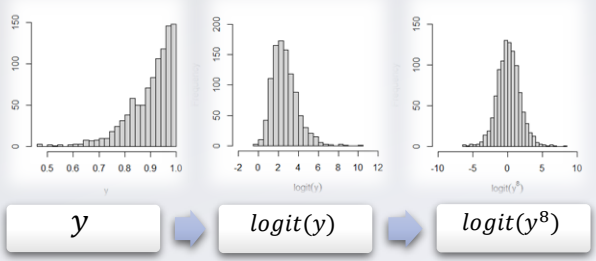
- GJS distributions are constructed from symmetric distributions assigned to the logit of Y.
- **Remark**: simple logit transformation is not able to bring common distributions of continuous fractions to symmetry.

- **Our approach**: **introduce a power parameter ($\lambda > 0$) in the logit transformation**

$$Z = \frac{1}{\sigma}\left[\text{logit}\left(Y^{\lambda}\right) - \text{logit}\left(\mu^{\lambda}\right)\right] \sim \text{S}(0, 1; r)$$

Y has a power logit (PL) distribution with parameters $0 < \mu < 1$, $\sigma > 0$, and $\lambda > 0$, representing the median, dispersion, and skewness of Y, respectively.

Power logit regression models



Definition and estimation

- $Y_1, \ldots, Y_n$ independent random variables with $Y_i \sim \mathrm{PL}(\mu_i, \sigma_i, \lambda; r)$.
- $r(\cdot)$ is the density generator function which may depend on an extra parameter (e.g. the degrees of freedom of the Student-t distribution).
- $d_1(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, $d_2(\sigma_i) = \boldsymbol{s}_i^\top \boldsymbol{\tau}$ where $d_1(\cdot)$ and $d_2(\cdot)$ are the link functions, $\boldsymbol{x}_i$ and $\boldsymbol{s}_i$ are the vectors of covariates, and $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are the parameter vectors.
- GJS regression models: $\lambda = 1$.
- The estimation process is based on a (penalized) maximum likelihood approach.

Residual analysis

Quantile residual	Deviance residual	Standardized residual
$r_i^q = \Phi^{-1}(R(\hat{z}_i))$	$r_i^d = \mathrm{sgn}(\hat{z}_i)\left\{2\log\left[\frac{r(0)}{r(\hat{z}_i^2)}\right]\right\}^{\frac{1}{2}}$	$r_i^s = \frac{\hat{z}_i}{\sqrt{\hat{\zeta}_r\left\{1-(\hat{d}_r\hat{\zeta}_r)^{-1}\hat{h}_{ii}\right\}}}$

Software implementation

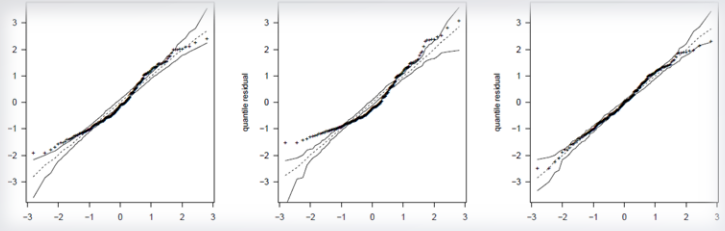
PLreg package

- Fitting both PL and GJS regression models
  - allows fixed λ
  - method for choosing the extra parameter
  - two types of estimators
- Diagnostic tools
  - three types of residuals
  - generalized leverage
  - diagnostic plots
- Influence analysis
  - local influence

Findings and conclusions

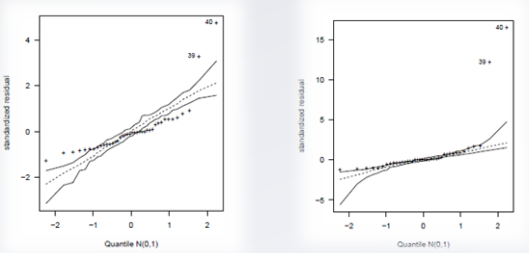
- Applications in real and simulated data showed some interesting features of the power logit models.

Flexibility



Beta model | GJS normal model | PL normal model

Deal with atypical observations



PL normal model | PL Student-t model

- The power logit regression models
  - are flexible,
  - allow fitting highly skewed data, and
  - deal well with outlying observations.
- Broad set of tools for likelihood inference and diagnostics.
- Software implementation.