clogitLasso: an R package for L1 penalized estimation of conditional logistic regression models
Marta Avalos, Hélène Pouyes

To cite this version:
Marta Avalos, Hélène Pouyes. clogitLasso: an R package for L1 penalized estimation of conditional logistic regression models. 1ères Rencontres R, Jul 2012, Bordeaux, France. <hal-00717505>
**clogitLasso: an R package for**

L^1 penalized estimation of conditional logistic regression models

**M. Avalos**\(^{a,b}\) and **H. Pouyes**\(^{a,c}\)

\(^{a}\)INSERM, ISPED, Centre INSERM U897–Epidemiologie–Biostatistique, F–33000 Bordeaux, France
\(^{b}\)Univ. Bordeaux, ISPED, Centre INSERM U897–Epidemiologie–Biostatistique, F–33000 Bordeaux, France
\(^{c}\)Univ. de Pau, Pau, France
marta.avalos@isped.u-bordeaux2.fr

**Keywords**: lasso, penalized conditional likelihood, matching, epidemiology.

The conditional logistic regression model is the standard tool for the analysis of epidemiological studies in which one or more cases (the event of interest), are individually matched with one or more controls (not showing the event). These situations arise, for example, in matched case–control studies and self–matched case–only studies (such as the case–crossover [1], the case–time–control [2] or the case–case–time–control [3] designs).

Usually, odds ratios are estimated by maximizing the conditional log–likelihood function and variable selection is performed by conventional manual or automatic selection procedures, such as stepwise. These techniques are, however, unsatisfactory in sparse, high-dimensional settings in which penalized methods, such as the lasso (least absolute shrinkage and selection operator) [4], have emerged as an alternative. In particular, the lasso and related methods have recently been adapted to conditional logistic regression [5].

The R package *clogitLasso* implements, for small to moderate sized samples (less than 3,000 observations), the algorithms discussed in [5], based on the stratified discrete-time Cox proportional hazards model and depending on the *penalized* package [6]. For large datasets, *clogitLasso* computes the highly efficient procedures proposed in [7, 8], based on an IRLS (iteratively reweighted least squares) algorithm [9] and depending on the *lassoshooting* package [10]. The most common situations that involve 1:1, 1:M and N:M matching are available.

The talk outlines the statistical methodology behind *clogitLasso* as well as its practical application by means of three real data examples arising from Epidemiology.

**References**

Future cases as present controls to adjust for exposure trend bias in case-only studies. *Epidemiology* **22**, 568-74.


