

# Conditional Logistic Regression Model

Stefanie Muff (stefanie.muff@uzh.ch), Johannes Signer, John Fieberg

July 16th 2018

## Parametrization

Some binomial sampling schemes in Biostatistics or Biology may result in what is called *matched case-control* data, which require a *conditional logistic regression* model. For the  $j^{\text{th}}$  observed binary response  $y_{nj}$  in stratum  $n$ , the model is given as

$$\text{Prob}(y_{nj} = 1 \mid \eta_{n\cdot}) = p_{nj} = \frac{\exp(\eta_{nj})}{\sum_i \exp(\eta_{ni})}, \quad y_{nj} \sim \text{Bern}(p_{nj}),$$

with linear predictor  $\eta_{nj}$  and success probability  $p_{nj}$ . The sum in the denominator is over all observations in the respective stratum. This model is a special case of a multinomial model, and as such it can be fitted by using a likelihood-equivalent reformulation as a Poisson model

$$\text{E}(y_{nj}) = \mu_{nj} = \exp(\alpha_n + \eta_{nj}), \quad y_{nj} \sim \text{Po}(\mu_{nj}),$$

with stratum-specific intercepts  $\alpha_n$ . If the number of strata is large, the explicit estimation of these intercepts can be circumvented by  $\alpha_n \sim \text{N}(0, \tau_\alpha)$  and fixing the precision  $\tau_\alpha$  at a very small value, e.g.  $10^{-6}$  or  $10^{-12}$ , which corresponds to a large variance. This mimicks a uniform distribution and ensures that the  $\alpha_n$  can be estimated freely instead of being shrunken towards 0.

## Hyperparameters

None.

## Specification

- family =Poisson
- To fix the variance at a large value the stratum-specific intercept  $\alpha_n$  use
  - model="iid"
  - hyper=list(theta = list(initial=log(1e-6),fixed=T))

## Example

The following example stems from a habitat selection study of 6 radio collared fishers (*Pekania pennanti*) (LaPoint et al. 2013), and was adapted from Signer et al. (2018). Outcomes with  $y = 1$  represent locations that were visited by fishers, and  $y = 0$  represents nearby locations that were not visited. Each visited location was matched to 2 nearby available locations, and together these 3 observations form a stratum (indicated by **stratum**). By design, only exactly one location can be visited in each stratum, thus these data need to be analyzed by a conditional logistic regression model. Covariates include sex (**sex**), land use (**landuse**, categorical covariate) and distance to the center of the habitat (**dist\_center**), with individual-dependent random slopes for **dist\_cent**. The 6 individuals are represented using **id** and **id1**. Shown is a reduced dataset with only 100 steps per individual and a sampling ratio of 1:2.

```

fisher.dat <- readRDS(system.file("demodata/data_fisher2.rds", package
= "INLA"))
fisher.dat$id1 <- fisher.dat$id
fisher.dat$dist_cent <- scale(fisher.dat$dist_cent)

formula.inla <- y ~ sex + landuse + dist_cent +
  f(stratum,model="iid",hyper=list(theta = list(initial=log(1e-6),fixed=T))) +
  f(id1,dist_cent, model="iid")

r.inla <- inla(formula.inla, family ="Poisson", data=fisher.dat)

```

## References

Muff, S., Signer, J. and Fieberg, J. (preprint) *Accounting for individual-specific variation in habitat selection studies: Efficient estimation using integrated nested Laplace approximations*

Signer, J., Fieberg, J. and Avgar, T. In press. *Animal Movement Tools (amt): R-Package for Managing Tracking Data and Conducting Habitat Selection Analyses*. Ecology and Evolution.

LaPoint, S., Gallery, P., Wikelski, M. and Kays, R. (2013) *Animal behavior, cost-based corridor models, and real corridors*. Landscape Ecology, 28, 1615–1630.