97

GLM for Generalized Linear Misleading? The need for a logistic function with estimated asymptotes to overcome inferential issues of the canonical logit link functions for Bernoulli GLMs

The study of the distribution of species, through presence absence data, is often based on canonical Bernoulli Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs). These models classically involve a logistic function with asymptotes fixed at zero and one. Although functions relaxing this assumption have already been proposed, the impacts of this assumption on inferential properties of estimators have not yet been studied.

Through a simulation study, we tested our hypothesis that in cases where the true asymptotes are different from zero and/or one and are approximately reached, the estimated parameters of the model could be strongly biased.

In the light of our results, three lessons can be drawn for the analysis of binary data: (i) when using canonical Bernoulli GLMs, the canonical link functions are suboptimal in cases where the asymptotes could be different from 0.0 and/or 1.0 and when they were nearly reached for a substantial proportion of data, inducing, in particular, a significant risk of underestimating the slope; (ii) when using canonical Bernoulli GAMs, the splines do not allow to overcome all the issues related to an improper link function and we can expect an underestimation of the magnitude of the relation at least when the asymptotes are different from 0.0 and/or 1.0, that they are almost reached, and when we have two continuous explanatory variables or one irregularly distributed continuous variable; (iii) finally, when using more advanced non-linear functions, the models with estimated asymptote(s) can not only be better than canonical form in terms of predictive capacity but also in terms of inference (estimation of the magnitude of the relationship).