Regression Models for Ordinal Data: Introducing *R*-package ordinal

Rune Haubo B Christensen[1,*]

1. Technical University of Denmark, Department of Informatics and Mathematical Modelling, Section for Mathematical Statistics, Richard Petersens Plads, Building 305, Room 122, DK-2800 Kgs. Lyngby, Denmark
*Contact author: rhbc@imm.dtu.dk

Keywords: Cumulative Link Models, Mixed Effects, Location-Scale, Partial Proportional Odds

Ordered categorical data, or simply *ordinal* data, are commonplace in scientific disciplines where humans are used as measurement instruments. Examples include school gradings, ratings of preference in consumer studies, degree of tumor involvement in MR images and animal fitness in field ecology. Cumulative link models (Agresti, 2002) are a powerful model class for such data since observations are treated rightfully as categorical, the ordered nature is exploited and the flexible regression framework allows in-depth analyses. A pertinent latent variable interpretation of cumulative link models is an important aspect in many applications in sensometrics, psychometrics and other social sciences. Cumulative link (mixed) models are implemented in functions `clm` and `clmm` in package **ordinal** (Christensen, 2011).

The **MASS** function `polr` is a popular implementation of basic cumulative link models taking its name from the proportional odds model—a cumulative link model with a logit link. `clm` and `clmm` extends `polr` in a number of ways by allowing for random effects, scale effects, nominal effects, flexible link functions and structured thresholds, further, several estimation methods are available including a fast Newton scheme. Mixed effects models are estimated with standard Gauss-Hermite quadrature, the highly accurate adaptive Gauss-Hermite quadrature or the flexible Laplace approximation accommodating nested as well as crossed random effect structures.

Collectively these options facilitate a fuller analysis of ordinal data. The model framework embrace location-scale models (McCullagh, 1980; Cox, 1995), allows for so-called partial proportional odds (Peterson and Harrell Jr., 1990), facilitates inference for the link function (Genter and Farewell, 1985) and allows assessment of linearity of the response scale. `profile` likelihood methods help visualize the likelihood function and provide accurate confidence intervals via a `confint` method. `dropterm`, `addterm` and `anova` methods facilitate model comparison. The implementation is primarily in *R*-code with computer intensive parts in *C*.

References

Agresti, A. (2002). *Categorical Data Analysis* (Second ed.). Wiley.

Christensen, R. H. B. (2011). ordinal—regression models for ordinal data. R package version 2010.12-15 http://www.cran.r-project.org/package=ordinal/.

Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in medicine 14*, pp. 1191–1203.

Genter, F. C. and V. T. Farewell (1985). Goodness-of-link testing in ordinal regression models. *The Canadian Journal of Statistics 13*(1), pp. 37–44.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B 42*, pp. 109–142.

Peterson, B. and F. E. Harrell Jr. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics 39*, pp. 205–217.