Journal of Statistical Software

*September 2025, Volume 114, Issue 9.* *doi: 10.18637/jss.v114.i09*

# GLMcat: An R Package for Generalized Linear Models for Categorical Responses

**Lorena León**
CIRAD, IMAG, Université de Montpellier

**Jean Peyhardi**
IMAG, CNRS, Université de Montpellier

**Catherine Trottier**
IMAG, CNRS, Université de Montpellier, AMIS Université Paul Valéry Montpellier 3

**Abstract**

In statistical modeling, there is a wide variety of generalized linear models for categorical response variables (nominal or ordinal responses); yet, there is no software embracing all these models together in a unique and generic framework. We propose and present **GLMcat**, an R package to estimate generalized linear models implemented under the unified specification $(r, F, Z)$ where $r$ represents the ratio of probabilities (reference, cumulative, adjacent, or sequential), $F$ the cumulative distribution function for the linkage, and $Z$ the design matrix. All classical models (and their variations) for categorical data can be written as an $(r, F, Z)$ triplet, thus, they can be fitted with **GLMcat**. The functions in the package are intuitive and user-friendly. For each of the three components, there are multiple alternatives from which the user should thoroughly select those that best address the objectives of the analysis. The main strengths of the **GLMcat** package are the possibility of choosing from a large number of link functions (defined by the composition of $F$ and $r$) and the simplicity for setting constraints in the linear prediction, either on the intercepts or on the slopes. This paper proposes a methodological and practical guide for the appropriate selection of a model considering the concordance between the nature of the data and the properties of the model.

*Keywords*: generalized linear model, categorical response, link function, cumulative models, sequential models, adjacent models, reference models.

## 1. Introduction

Regression models for categorical responses have emerged in various disciplines and under different names. The underlying structures of such models may be closely related (or even the same) but they are often perceived as fundamentally different. The intrinsic differences among the classical models concern the assumed link function and the specification of the linear predictor. Once the linear predictor is defined, the question to address concerns the most appropriate link function. The selected link function should reflect the nature of the response variable; for categorical responses, a broad distinction is made on the basis of the scale itself, being either nominal or ordinal.

For ordinal responses, there are three families of generalized linear models (GLMs): the cumulative, the sequential, and the adjacent models. The family of cumulative models (simply known as ordinal regression models) is the most popular. This family includes the odds proportional logit model (McCullagh 1980) which has been the most widely used model for ordinal data. Sequential models have initially been discussed by authors including Fienberg (1980), Armstrong and Sloan (1989) and Tutz (1991). The most iconic model within this family is the proportional hazard model which was originally developed by Cox (1972) for continuous responses. More recently, Fahrmeir and Tutz (2001) briefly proposed an extension of the adjacent logit model (Goodman 1983; Agresti 1989) that allows substituting the logistic distribution function by any other CDF. In this light, Peyhardi, Trottier, and Guédon (2015) detailed the estimation of such models and named them the family of adjacent models. Adjacent models have been widely adopted in item response theory, a widespread model within this framework is the polytomous Rasch model (Rasch 1961; Andersen 1995) which is an adjacent logit model with a specific form for the linear predictor. In contrast to ordinal responses, for which there exists a variety of GLMs, until recently, the only option for nominal responses (within the context of GLMs) was the multinomial logit (MNL) model, introduced by Luce (1959). To fill this gap, Peyhardi *et al.* (2015) generalized the structure of the MNL allowing the use of several cumulative distribution functions (CDFs) as alternatives to the logistic CDF, the resulting set of models is referred to as the family of reference models.

Notwithstanding the wide set of model options, the use of appropriate models for categorical responses seems to be rather limited in the literature (Ananth and Kleinbaum 1997; Liddell and Kruschke 2018). Besides, on the rare occasions when they are employed, there is often no consistency between the response variable characteristics and the model's assumptions. For instance, ordinal responses have been frequently treated by researchers as standard nominal or metric problems. Another usual and inaccurate approach is to dichotomize categorical responses with the aim of using the standard logistic or probit regression models (Sankey and Weissfeld 1998). These pitfalls can lead to non-optimal solutions and thus to erroneous statistical inferences (see Liddell and Kruschke 2018; Scott, Goldberg, and Mayo 1997; Gutiérrez, Pérez-Ortiz, Sánchez-Monedero, Fernández-Navarro, and Hervás-Martínez 2016, for further details). Despite the current availability of statistical software to fit models which take full advantage of the ordinal nature of the response, the described poor practices are still common. As noted by Mellenbergh (1995), one possible cause may be linked to the fact that the literature for ordinal responses does not provide much support for preferring one family of models over another. In addition, we suspect that the lack of homogeneity that characterizes the literature and the software solutions for this subject may result confusing and overwhelming. Hence, it is not surprising that, risking a loss of accuracy and interpretability, the user might opt for the most popularly used models.

In R (R Core Team 2025) there is a variety of packages to fit categorical responses, however, most of them only cover one or a few of the types of models. For instance, the function `multinom()` of the package **nnet** (Ripley and Venables 2025; Venables and Ripley 2002) fits the MNL via neural networks. For ordinal responses, the functions `polr()` of the package

**MASS** (Ripley, Venables, Bates, Hornik, Gebhardt, and Firth 2025; Venables and Ripley 2002) and `omr()` from the **rms** (Harrell 2025) package are often used to fit the odds proportional model. Few packages are aimed to fit a whole family of models for categorical responses, one of them is the **tram** (Hothorn, Siegfried, and Kook 2025; Hothorn 2020) package, which by means of the `Polr()` function allows for stratification, censoring and truncation in the response of cumulative models. The **ordinal** package (Christensen 2024) is another option to fit the family of cumulative models, it includes a comprehensive implementation of this class of models offering great flexibility, notably in the specification of the linear predictor. To our knowledge, only the **brms** (Bürkner 2017) and the **ordinalNet** (Wurm, Rathouz, and Hanlon 2025) packages consider the three families of ordinal models: cumulative, sequential, and adjacent. The **VGAM** package also consider these families but their adjacent ratio seems to be valid only for the logistic distribution. It enables however to implement its own cumulative distribution function. Let us note that categorical regression is just a part of the large **VGAM** package's possibilities (Goodman $R \times C$ association models, Bradley-Terry models, genetic models, etc.). None of the aforementioned packages encloses the four model families for categorical responses and most of them have some limitations in terms of adding constraints to the design, or in the availability of the CDF that one can use as part of the link function. These gaps also exist in commercial statistical software like SAS (SAS Institute Inc. 2020), Stata (StataCorp 2015), and SPSS (IBM Corporation 2017). An additional problem of the commercial packages is that those use different techniques (which are not strictly equivalent) to fit the models. As a consequence, different estimations might be obtained when using different software, even though the same theoretical model is specified. For instance, Liu (2009) reported some differences in the estimation of an odds proportional model using the functions `PROC LOGISTIC` in SAS, `ologit` in Stata, and `PLUM` in SPSS.

Despite the diverse origins, names, applications, and implementations of the above-mentioned models, they all share a common structure that was fully described by Peyhardi *et al.* (2015). The authors introduced a unified specification of GLMs for categorical responses that encompasses the four families of models based on a decomposition of the link function. They introduced the notation $(r, F, Z)$ for this decomposition where: $r$ is the ratio that characterizes the ordering type of the response variable, $F$ is the CDF of the link function, and $Z$ is the design matrix where the form of the linear predictor is specified. The comprehensive description of the taxonomy of the GLMs for categorical data given by the $(r, F, Z)$ decomposition exposes the fundamentals, relations, and equivalences of the families of models. Furthermore, the possible extensions for each model family become evident and can be easily implemented. These extensions are obtained by structuring the design matrix (for intercepts and slopes), as well as broadening the spectrum of CDFs. The logistic CDF is usually chosen by default since other links, such as probit, may appear very similar. Nevertheless, we present in Section 4 an example to highlight how much the use of the Student link improves notably the model accuracy by detecting noisy variables in this case. Exploring different distribution functions enables to uncover hidden patterns, detect outliers and improve the overall understanding of the underlying data.

We implemented the $(r, F, Z)$ methodology in the **GLMcat** (León, Peyhardi, and Trottier 2025) package developed for R, available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=GLMcat. Our purpose is to provide an alternative that covers all the classical models for categorical responses and which gives room to extend them through different components. The package supports a wide range of CDFs and allows to adapt the linear predictor at any desired extent. In consideration of all these possibilities, we intend to guide the user in the identification of the most pertinent combination of the ratio, the CDF, and the design matrix, highlighting the limitations or advantages of the resulting $(r, F, Z)$ model. In the **GLMcat** package there are two main functions: `glmcat()`, which covers the four families of models for categorical responses, and `discrete_cm()`, which extends the family of reference models to take into account explanatory variables that depend on response categories (useful for the discrete choice model).

The content of the paper is presented in three main sections. In Section 2, we recall the unified specification of GLMs through the $(r, F, Z)$ triplet and we illustrate its implementation in **GLMcat**. We also describe in detail each of the three components as well as the possible extensions for them. In Section 3, we aim to characterize the different families of models for ordinal responses by outlying a series of properties inherent to each of them. We emphasize the importance of identifying the model that makes the appropriate assumptions in light of the nature of the response variable and the goals of the analysis. In Section 4, we revisit the family of reference models in the framework of discrete choice models (Bouscasse, Joly, and Peyhardi 2019; Peyhardi 2020). We motivate the use of this family of models presenting its strengths in contrast to existing alternatives. The model fitting by means of the **GLMcat** package is illustrated in all the sections using different datasets and its computational implementation is described in Section 2.5.

## 2. Unified specification of GLMs for categorical data

Consider the regression context where the response $Y$ is a categorical variable (with $J \geq 2$ categories). The aim is to model the effect on $Y$ of a given set of $q$ explanatory variables $\boldsymbol{x} = (x_1, \ldots, x_q)$ defined in a general form of dimension $p$ with $p \geq q$ (i.e., categorical variables are represented by indicator vectors). In the following, we will sometimes use the univariate notation $\{Y = j\}$ or, equivalently, the indicator vector notation $\boldsymbol{Y} = (Y_1, \ldots, Y_{J-1})$ with 1 in position $j$ and 0 elsewhere. Note that $\{Y = J\}$ would correspond to $\boldsymbol{Y} = (0, \ldots, 0)$. For convenience, models are presented at the individual level, thus, the subscript $i \in \{1, \ldots, n\}$ is not mentioned. A GLM for categorical response can be decomposed into three parts:

1. the random component which accounts for the conditional distribution of the response variable given the set of the explanatory variables. In the framework of categorical response variables, $Y$ follows the multinomial distribution

$$\boldsymbol{Y}|\boldsymbol{x} \sim \mathcal{M}(1, \boldsymbol{\pi}(\boldsymbol{x}))$$

with $\boldsymbol{\pi} = (\pi_1(\boldsymbol{x}), \ldots, \pi_{J-1}(\boldsymbol{x})) \in \Delta$ where $\Delta = \{\boldsymbol{\pi} \in (0,1)^{J-1} : \sum_{j=1}^{J-1} \pi_j < 1\}$. This is a generalization of the Bernoulli distribution (obtained when $J = 2$). The probability mass function written in terms of $\boldsymbol{y}$ is then

$$f(\boldsymbol{y}; \boldsymbol{\pi}) = \left( \prod_{j=1}^{J-1} \pi_j^{y_j} \right) \left( 1 - \sum_{j=1}^{J-1} \pi_j \right)^{1 - \sum_{j=1}^{J-1} y_j}.$$

Its natural parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{J-1})^\top$ is defined by

$$\boldsymbol{\theta} = \left( \log\left( \frac{\pi_1}{1 - \sum_{j=1}^{J-1} \pi_j} \right), \ldots, \log\left( \frac{\pi_{J-1}}{1 - \sum_{j=1}^{J-1} \pi_j} \right) \right)^\top,$$

and

$$b(\boldsymbol{\theta}) = \log\left(1 + \sum_{j=1}^{J-1} e^{\theta_j}\right).$$

Based on the above decomposition, the probability mass function can be simply written as

$$f(\boldsymbol{y};\boldsymbol{\theta}) = \exp\{\boldsymbol{y}^\top\boldsymbol{\theta} - b(\boldsymbol{\theta})\}.$$

Using the weight $\omega = 1$, the dispersion parameter $\phi = 1$ and the null function $c(\boldsymbol{y},\phi) = 0$, we see that this distribution function belongs to the exponential family of dimension $K = J - 1$.

2. The systematic component which is determined by the linear predictor $\boldsymbol{\eta} = (\eta_1,\dots,\eta_{J-1})$. For each category $j$, the linear predictor has the form $\eta_j = \alpha_j + \boldsymbol{x}^\top\boldsymbol{\delta}_j$, where $\alpha_j \in \mathbb{R}$ is the intercept and $\boldsymbol{\delta}_j \in \mathbb{R}^p$ is the vector of slopes. Considering the parameter vector as $\boldsymbol{\beta} = (\alpha_1,\dots,\alpha_{J-1},\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_{J-1})$, the linear predictor can be written as the product

$$\boldsymbol{\eta} = Z\boldsymbol{\beta},$$

where $Z$ denotes the design matrix composed of repetitions of 1 and $\boldsymbol{x}^\top$ (see Section 2.3 for some examples of design matrices).

3. The link function $g$ which relates the conditional expectation of the response variable $\boldsymbol{\pi} = \mathsf{E}[\boldsymbol{Y}|\boldsymbol{x}]$ and the linear predictor $\boldsymbol{\eta}$. The equality $g(\boldsymbol{\pi}) = \boldsymbol{\eta}$ corresponds to the $J-1$ equations $g_j(\pi) = \eta_j$.

Peyhardi *et al.* (2015) showed that all the classical link functions can be decomposed through the unified specification

$$g_j = F^{-1} \circ r_j \quad \Leftrightarrow \quad r_j(\boldsymbol{\pi}) = F(\eta_j) \qquad j = 1,\dots,J-1, \tag{1}$$

where $F$ is a continuous and strictly increasing CDF, and $\boldsymbol{r} = (r_1,\dots,r_{J-1})$ is a map from the simplex $\Delta$ to the open hypercube $(0,1)^{J-1}$. We call $\boldsymbol{r}$ the ratio even if for the cumulative family it is not defined as a proper ratio of probabilities. Let see in the next section an expression of the probabilities $\boldsymbol{\pi}$ from the predictors $\boldsymbol{\eta}$ according to each ratio type to facilitate the interpretation. The authors introduced the notation $(r, F, Z)$ with which any classical GLM for categorical responses can be fully described. Throughout this paper and in this framework, we interchangeably use the terms $(r, F, Z)$ and GLM.

The **GLMcat** package is designed based on the $(r, F, Z)$ decomposition. To facilitate the user experience, instead of calling a specific function for each family of models (determined by the ratio), we implemented a single function: `glmcat()`, with which any $(r, F, Z)$ model can be fitted. In the following, we will describe more closely the components $r$, $F$, and $Z$ and their modalities.

## 2.1. Ratio of probabilities $r$

In models for categorical responses, the linear predictor $\boldsymbol{\eta}$ is not directly related to the expectation $\boldsymbol{\pi}$ but to a particular transformation $\boldsymbol{r}$ of the vector $\boldsymbol{\pi}$.

| | Cumulative | Sequential | Adjacent | Reference |
|---|---|---|---|---|
| $r_j(\boldsymbol{\pi})$ | $\pi_1 + \ldots + \pi_j$ | $\dfrac{\pi_j}{\pi_j + \ldots + \pi_J}$ | $\dfrac{\pi_j}{\pi_j + \pi_{j+1}}$ | $\dfrac{\pi_j}{\pi_j + \pi_J}$ |
| $Y$ | | ordinal | | nominal |

Table 1: Four ratios $r_j(\boldsymbol{\pi})$ of GLMs for categorical responses ($j = 1, \ldots, J-1$).

The ratios for categorical responses are defined in Table 1. The cumulative ratio of category $j$ is the result of the cumulated probabilities of the precedent categories:

$$\pi_1 + \cdots + \pi_j = F(\eta_j) \quad \Leftrightarrow \quad \pi_j = F(\eta_j) - F(\eta_{j-1}),$$

with the conventions $\eta_0 = -\infty$, $\eta_J = +\infty$ and the ordering assumption $\eta_0 < \eta_1 < \cdots < \eta_{J-1} < \eta_J$. For the sequential ratio each category $j$ is compared to its following categories $j+1, \ldots, J$

$$\frac{\pi_j}{\pi_j + \ldots + \pi_J} = F(\eta_j) \quad \Leftrightarrow \quad \pi_j = F(\eta_j) \prod_{k=1}^{j-1} (1 - F(\eta_k)),$$

with the convention $\prod_{k=1}^{0}(1 - F(\eta_k)) = 1$. For the adjacent ratio each category $j$ is compared to its adjacent category

$$\frac{\pi_j}{\pi_j + \pi_{j+1}} = F(\eta_j) \quad \Leftrightarrow \quad \pi_j = \frac{\prod_{k=j}^{J-1} \frac{F(\eta_k)}{1-F(\eta_k)}}{1 + \sum_{l=1}^{J-1} \prod_{k=l}^{J-1} \frac{F(\eta_k)}{1-F(\eta_k)}}$$

The adjacent, cumulative, and sequential ratios rely on an ordering assumption among categories. On the contrary, the reference ratio relates each category $j$ to a reference category ($J$ by convention)

$$\frac{\pi_j}{\pi_j + \pi_J} = F(\eta_j) \quad \Leftrightarrow \quad \pi_j = \frac{\frac{F(\eta_j)}{1-F(\eta_j)}}{1 + \sum_{k=1}^{J-1} \frac{F(\eta_k)}{1-F(\eta_k)}}$$

No ordering assumption among categories is needed in this case. The reference ratio is therefore devoted to nominal responses. For each ratio, the interpretation of an explanatory effect is different. An increase of the predictor $\eta_j$ positively impacts the ratio of probabilities $r_j(\boldsymbol{\pi})$ since the CDF is strictly increasing. Note that the ratio $r_j(\boldsymbol{\pi})$ is a probability itself equal to $\mathsf{P}(Y \le j)$, $\mathsf{P}(Y = j | Y \ge j)$, $\mathsf{P}(Y = j | Y \in \{j, j+1\})$ and $\mathsf{P}(Y = j | Y \in \{j, J\})$ respectively for the cumulative, sequential, adjacent and reference ratio. Therefore it is more convenient to interpret an explanatory effect, using the left part of each equation. The ratios are the essential units from which a family of models is defined. For this reason, we named the model families according to their corresponding ratio. In **GLMcat** we cover the four ratios described above, whereby all known GLMs for categorical responses are within reach by handling one single package. The ratio should be specified in the `glmcat()` function as a string in the argument `ratio`. If no ratio is specified, the reference ratio will be used by default.

## 2.2. Cumulative distribution function $F$

The link function in the binary regression framework accounts only for the CDFs that link the expected value of the response to the linear predictor of the model. Based on the decomposition presented in Equation 1, it is evident that for the $(r, F, Z)$ models the CDF is just one part of the link function, which along with the ratio, characterizes the relation between $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$. Remark that the CDF is assumed to be differentiable and strictly increasing. The differentiability is necessary for the Fisher's scoring algorithm (or Newton Raphson's algorithm) computation. The condition of strict increase is necessary for parameter interpretation since the coefficients $\delta_{j,k}$ give the signs of the partial effects of the corresponding explanatory variable $x_k$ on the probability $r_j(\boldsymbol{\pi})$.

The distinction between symmetric and asymmetric CDFs has an impact on the properties of the models as it will be demonstrated later. Moreover, the choice of the distribution might markedly improve the model fit. In literature, there are different recommendations to choose the CDF of a GLM, although the logistic distribution is the most widely used. The choice is often related to disciplines or fields. For instance, economists tend to favor the normal distribution due to its association with the utility notion; the Gumbel distribution is popular in survival and hazard analysis, since it can appropriately model the occurrence of events. The aforementioned CDFs are available in many packages. **GLMcat** proposes, in addition, some less popular alternatives such as the Cauchy, the Gompertz, the Laplace, the Student, and the non-central Student CDFs. In particular, the Student CDF has proven to be a robust alternative for regression models (see Lange, Little, and Taylor 1989; Peyhardi 2020) and can be considered as a family of functions given that the shape of the CDF varies according to $\nu$, the degrees of freedom. All of the CDFs presented in Table 2 are available in **GLMcat** and should be specified by its name as a string in the argument `cdf`. For the particular case of Student CDF, as the degrees of freedom are to be specified, the user must enter a list of the form `list(cdf = "student", df = 7)`. If the CDF is not declared, the logistic distribution is used by default. In the following, the set of cumulative distribution functions will be denoted by $\mathfrak{F}$.

*Normalization of parameter estimates*

The parameters of the models with different CDFs are not comparable since they refer to specific means and variances (Tutz 2011). Often, the parameter estimates will turn out to be different even if there is not an apparent discrepancy in terms of goodness-of-fit. Tutz (2011) illustrated an alternative to standardize the parameters of a binary regression model:

$$\tilde{\alpha} = \frac{\alpha - \mathsf{E}(\varepsilon)}{\sqrt{\mathsf{VAR}(\varepsilon)}}, \quad \tilde{\delta} = \frac{\delta}{\sqrt{\mathsf{VAR}(\varepsilon)}}, \quad \text{where } \varepsilon \sim F.$$

Note that this approach is not suitable when using a CDF with undefined mean or variance as it is the case of the Student distribution (whose mean and variance are not defined when $\nu \leq 1$, and $\nu \leq 2$, respectively). A propagated approach in econometrics that solves this problem is to consider the average partial effect of the variable $x_k$ on $\pi_j$ as the scale factor, i.e., $\partial\pi_j(\boldsymbol{x})/\partial x_k$ (Wooldridge 2012). If $x_k$ is a continuous variable, its partial effect on $\pi_j(\boldsymbol{x})$ is obtained from the partial derivative:

$$\frac{\partial \pi_j(\boldsymbol{x})}{\partial x_k} = \frac{\partial F}{\partial \eta_j}\frac{\partial \pi_j}{\partial r_j}\hat{\delta}_{j,k}\,. \tag{2}$$

The average partial effect of $x_k$ on $\pi_j$ is then given by the mean value over the individuals. The downsides of this method are that the scale factor depends on the input data and that it is only valid for continuous explanatory variables. Note that Equation 2 results in $f(\eta_j)\hat{\delta}_k$ for

| Distribution | $F(\eta)$ | Shape |
|---|---|---|
| Logistic | $\frac{1}{1+\exp(-\eta)}$ | Symmetric |
| Normal | $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\eta}\exp\left(-\frac{x^2}{2}\right)dx$ | Symmetric |
| Laplace | $\begin{cases}\frac{1}{2}\exp(\eta) & \text{if } \eta < 0,\\ 1-\frac{1}{2}\exp(-\eta) & \text{if } \eta \geq 0\end{cases}$ | Symmetric |
| Cauchy | $\frac{1}{2}+\frac{1}{\pi}\arctan(\eta)$ | Symmetric |
| Student($\nu$) | $\frac{1}{2}+\eta\Gamma\left(\frac{\nu+1}{2}\right)$ | Symmetric |
| Non-central $t$ ($\nu$, $\mu$) | $\begin{cases}F_{\nu,\mu}(\eta) & \text{if } \eta \geq 0,^{(1)}\\ 1-F_{\nu,-\mu}(\eta) & \text{if } \eta < 0\end{cases}$ | Left skewed if $\mu < 0$,<br>Symmetric if $\mu = 0$,<br>Right skewed if $\mu > 0$. |
| Gompertz | $1-\exp(-\exp(\eta))$ | Left skewed |
| Gumbel | $\exp(-\exp(-\eta))$ | Right skewed |

Table 2: List of the CDFs available in **GLMcat** to use as part of the link function for GLMs. $^{(1)}$ Refer to Appendix A for the complete form of $F_{\nu,\mu}$.

the binomial regression. In this case, if $f$ is a symmetric probability density function (PDF) around zero, the largest effect occurs when $\eta = 0$. For instance, for the normal PDF, this will be at $f(0) \approx 0.4$ and for the logistic PDF at $f(0) = 0.25$. A simple approach to make the magnitudes of those two CDFs roughly the same is to multiply the probit estimates by $0.4/0.25 = 1.6$ or to multiply the logit estimates by $0.25/0.4 = 0.625$.

Bouscasse *et al.* (2019) proposed a normalization of parameter estimates via the location parameter $m$ and the scale parameter $s$ of the CDF $F$. Two real points $a$ and $b$ are chosen such that all CDFs in $\mathfrak{F}$ have the same values for $F(a)$ and $F(b)$. It is imposed that $F(0) = 1/2$ to preserve the interpretability of the intercepts. Note that the reference, the adjacent and the sequential ratios satisfy this condition. To illustrate this, assume $\delta_j = 0$ so that the linear predictor only depends on the intercept, i.e., $\eta_j = \alpha_j$. For reference models we obtain $\pi_j/\pi_J = F(\alpha_j)/(1-F(\alpha_j))$, hence, if the intercept is null, and setting $F(0) = 1/2$, it is evident that the probabilities $\pi_j$ and $\pi_J$ are equal and so are all the elements in $\boldsymbol{\pi}$. This equality is also valid for the adjacent ratio but neither for sequential nor cumulative. For the sequential family, we can find that the probabilities will correspond to $\pi_j = (1/2)^j$ for $j = 1, \dots, J-1$, and $\pi_J = (1/2)^{J-1}$. Conversely, for the cumulative ratio, the intercepts must be strictly ordered and cannot be all equal to zero. Therefore for this model, the constraint $F(0) = 1/2$ is not necessary. Remark that the condition $F(0) = 1/2$ is already satisfied for the symmetric distributions and has to be imposed for asymmetric CDFs. The logistic distribution is proposed as the reference CDF since it is part of the canonical link function. Thus, the second point is given by $F(b) = e^b/(1+e^b)$. The authors suggested to use the quantile of the logistic distribution such that $b = q_p$ for some $p > 1/2$. The normalized space is then $\mathfrak{F}_{q_p} = \{F \in \mathfrak{F} : F(0) = 1/2,\ F(q_p) = p\}$. We have that $F_{m_0,s_0} \in \mathfrak{F}_{q_p}$ if

$$\begin{cases} m_0 = \dfrac{F^{-1}(1/2) \cdot q_p}{F^{-1}(p) - F^{-1}(1/2)} \\ s_0 = \dfrac{q_p}{F^{-1}(p) - F^{-1}(1/2)}. \end{cases}$$

The normalized parameters using the above approach are: $\alpha'_j = m_0 + \alpha_j s_0$ and $\delta'_j = \delta_j s_0$ for $j = 1, \dots, J-1$. We implemented this normalization since it works for any number of categories, for any type of explanatory variables, and because it does not depend on the dataset. In the functions of **GLMcat**, the normalization using the quantile $q_{0.95}$ (which can be considered as the standard case, $q_{0.95} \approx 2.94$) is obtained with the argument `normalization = 0.95`. The `summary()` function returns the transformed parameters when specifying the argument `normalized = TRUE`. An example of the normalization of parameters is illustrated in Section 4.

## 2.3. Design matrix $Z$

In a linear predictor, one can define constraints to model the effect of the explanatory variables on the categorical response. Commonly, these constraints have been imposed only on the slopes and not much attention has been given to the intercepts. In **GLMcat**, we divide the design matrix into two blocks: $I$ to control the intercepts and $S$ to control the slopes, i.e.,

$$Z = \left( \begin{array}{c|c} I & S \end{array} \right).$$

The design matrix can be fully customized using this decomposition. By default for the reference family of models, the **GLMcat** package proposes a complete (also known as non-proportional effects or category-specific effects) design without any constraint on the effects, i.e., $Z_c = (I_c|S_c)$. This matrix is of dimension $(J-1) \times (J-1)(1+p)$, and has the form:

$$Z_c = \left( \begin{array}{ccc|ccc} 1 & & & \boldsymbol{x}^\top & & \\ & \ddots & & & \ddots & \\ & & 1 & & & \boldsymbol{x}^\top \end{array} \right).$$

*Slope design matrix $S$:*

the most common constraint is to impose the effects of the explanatory variables to be constant across the response categories, thus, it is assumed the existence of a single global effect for each explanatory variable. This constraint is known as the parallelism or proportional assumption, and the user should verify its validity before using it (Harrell 2015). The slope matrix associated with the parallel design is of dimension $(J-1) \times p$ and has the form:

$$S_p = \begin{pmatrix} \boldsymbol{x}^\top \\ \vdots \\ \boldsymbol{x}^\top \end{pmatrix}.$$

A more flexible framework is to consider both kinds of effects, complete and parallel, the resulting design is known as partial parallel. The following slope matrix of dimension $(J -$

$1) \times ((J-1)p_1 + p_2)$ represents the design for $p_1$ explanatory variables $\boldsymbol{x}^1 = (x_1, \ldots, x_{p_1})$ with complete design effects, and $p_2$ explanatory variables $\boldsymbol{x}^2 = (x_{p_1+1}, \ldots, x_{p_1+p_2})$ with parallel effects:

$$S_{cp} = \begin{pmatrix} \boldsymbol{x}^{1\top} & & & \boldsymbol{x}^{2\top} \\ & \ddots & & \vdots \\ & & \boldsymbol{x}^{1\top} & \boldsymbol{x}^{2\top} \end{pmatrix}.$$

The `glmcat()` function assumes by default a parallel design for the cumulative, sequential, and adjacent ratios. If all explanatory variables are to be set with the complete design, one should simply specify `parallel = FALSE`. If the user opts for the partial parallel design, the variables with a parallel effect must be specified in a string vector in the argument `parallel`.

In Section 4, we further explore the design matrix particularly for nominal response variables for which the function `discrete_cm()` allows to specify a particular response category on which the explanatory variable(s) is expected to have an effect.

*Intercept design matrix $I$:*

if a single intercept is expected in the linear predictor, the intercept matrix $I_p$ is simply the vector $\mathbb{1}$ of size $J-1$. The design matrix $I_p$ is obtained by specifying the string `"(Intercept)"` in the argument `parallel`. Remark that in most categorical regression models, even for the minimal model, intercepts are assumed different for each category. Thus the parallel design is used to designate the design matrix $Z_p = (I_c|S_p)$. The *complete* design $Z_c = (I_c|S_c)$ and the *parallel* design $Z_p = (I_c|S_p)$ are sufficient to define all the classical models (see Table 3). Nevertheless, the constraints on $I$ can be further explored.

Christensen (2024) presented some constraints on the intercept for the cumulative models. For instance, if the distances between the adjacent intercepts are required to be the same for all pairs $(j, j+1)$, we can write the intercepts as $\alpha_j = \alpha_1 + (j-1)\theta$ for $j = 1, \ldots, J-1$. In that case, $\alpha_1$ corresponds to the first intercept and $\theta$ to the constant distance between intercepts. This restriction implies that, regardless of the number of categories, only two parameters must be estimated. The associated design matrix is of dimension $(J-1) \times 2$ and has the form:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & J-2 \end{pmatrix}.$$

Another form is given when the intercepts are symmetric around zero, i.e., the categories are supposed to be equally distant from the central category/categories. For an even and for an odd number of response categories, the dimension of the intercept matrix is $(J-1) \times J/2$ and $(J-1) \times (J+1)/2$, respectively, and the intercepts and their design matrices are respectively written as:

$$\alpha_j = \begin{cases} \theta_{J/2} - \theta_j & \text{if } j < J/2, \\ \theta_{J/2} & \text{if } j = J/2, \\ \theta_{J/2} + \theta_{J-j} & \text{if } j > J/2, \end{cases} \quad \text{and} \quad \alpha_j = \begin{cases} \theta_j & \text{if } j < J/2, \\ \theta_{J-j} & \text{otherwise,} \end{cases}$$

$$\begin{pmatrix} 1 & -1 & & \\ \vdots & & \ddots & \\ & & & -1 \\ & 0 & \cdots & 0 \\ & & & 1 \\ & & \iddots & \\ 1 & 1 & & \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 & & \\ \vdots & \ddots & \\ & & -1 \\ & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}.$$

The constraints on the intercepts are only available for the cumulative ratio and should be specified through the argument `threshold = "equidistant"` or `threshold = "symmetric"`. The computational instability that is frequently found in the cumulative models can be alleviated with the use of this constraint given that the number of parameters is reduced. An example of the use of the structured intercepts for cumulative models is presented by Reinhard, Rutrecht, Hengstenberg, Tutulmaz, Geissler, Hecht, and Muttray (2017). In the following, the set of design matrices will be denoted by $\mathfrak{Z}$.

### 2.4. $(r, F, Z)$ genericity

A large number of models for categorical responses have been proposed in the literature. Depending on the scientific context, some of these models can be differently named despite having the same formulation. In consequence, the relationships among them are often unrecognized. Earlier in this paper, we mentioned that any GLM for categorical responses can be written as the triplet $(r, F, Z)$. In Table 3 we present some of the best-known models in their original formulation and decomposed into the three components $r$, $F$, and $Z$. For categorical responses, the $(r, F, Z)$ specification enlarges the number of possible models to consider. Furthermore, it eases the comparison between them as we are going to demonstrate in the following.

### 2.5. Estimation and computational implementation

The **GLMcat** package can be installed within R (R Core Team 2025) using the line of code: `install.packages("GLMcat")`. The standard arguments `formula` and `data` are already known from the `lm()` and `glm()` functions from the **stats** package. The key difference is that in the `glmcat()` function, the link of the model must be specified through the two arguments `ratio` and `cdf`. In **GLMcat**, the response (categorical) variable must be defined as a factor or an ordered factor. The user can specify/change the order of the factor levels by means of the `ordered()` function. Alternatively, and for ease of use, one can indicate the order as a character vector in the argument `categories_order`. An example of the syntax of the `glmcat()` function for an ordinal response is

```
R> glmcat(formula = Level ~ Age, data = DisturbedDreams,
+   categories_order = c("Not.severe", "Severe.1", "Severe.2",
+     "Very.severe"), ratio = "adjacent", cdf = "gompertz")
```

For non-ordered response variables, the user must use the reference ratio, for which by default, the reference category is set to be the last level of the response factor variable. The user can also specify manually the reference category in the argument `ref_category`. Assuming that there is no order in the response of the `DisturbedDreams` dataset, the model would be fitted as follows:

| | |
|---|---|
| The multinomial logit model $\mathsf{P}(Y=j)=\dfrac{\exp(\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}_j)}{1+\sum_{k=1}^{J-1}\exp(\alpha_k+\boldsymbol{x}^t\boldsymbol{\delta}_k)}$ | $(\mathit{reference}, \mathit{logistic}, \mathit{complete})$ $\dfrac{\pi_j}{\pi_j+\pi_J}=\dfrac{1}{1+\exp(-\alpha_j-\boldsymbol{x}^t\boldsymbol{\delta}_j)}$ |
| The odds proportional logit model $\ln\left\{\dfrac{\mathsf{P}(Y\le j)}{1-\mathsf{P}(Y\le j)}\right\}=\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}$ | $(\mathit{cumulative}, \mathit{logistic}, \mathit{parallel})$ $\pi_1+\cdots+\pi_j=\dfrac{1}{1+\exp(-\alpha_j-\boldsymbol{x}^t\boldsymbol{\delta})}$ |
| The proportional hazard model $\ln\{-\ln\{\mathsf{P}(Y>j\|Y\ge j)\}=\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}$ | $(\mathit{sequential}, \mathit{Gompertz}, \mathit{parallel})$ $\dfrac{\pi_j}{\pi_j+\cdots+\pi_J}=1-\exp(-\exp(\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}))$ |
| The adjacent logit model $\ln\left\{\dfrac{\mathsf{P}(Y=j)}{\mathsf{P}(Y=j+1)}\right\}=\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}_j$ | $(\mathit{adjacent}, \mathit{logistic}, \mathit{complete})$ $\dfrac{\pi_j}{\pi_j+\pi_{j+1}}=\dfrac{1}{1+\exp(-\alpha_j-\boldsymbol{x}^t\boldsymbol{\delta}_j)}$ |
| The continuation ratio logit model $\ln\left\{\dfrac{\mathsf{P}(Y=j)}{\mathsf{P}(Y>j+1)}\right\}=\alpha_j+\boldsymbol{x}^t\boldsymbol{\delta}_j$ | $(\mathit{sequential}, \mathit{logistic}, \mathit{complete})$ $\dfrac{\pi_j}{\pi_j+\cdots+\pi_J}=\dfrac{1}{1+\exp(-\alpha_j-\boldsymbol{x}^t\boldsymbol{\delta}_j)}$ |

Table 3: $(r, F, Z)$ specification of some classical GLMs for categorical responses.

```
R> data("DisturbedDreams", package = "GLMcat")
R> glmcat(formula = Level ~ Age, data = DisturbedDreams,
+    ref_category = "Very.severe", ratio = "reference", cdf = "gompertz")
```

The object generated by the `glmcat()` function is compatible with the usual generic methods: `coef()` for the parameter estimates $\hat{\boldsymbol{\beta}}$ and `confint()` for their confidence intervals, `logLik()` for the log-likelihood, `nobs()` for the number of observations $n$, `predict()` to obtain $\hat{\boldsymbol{\eta}}$ (if `type = "linear.predictor"`) or $\hat{\boldsymbol{\pi}}$ (if `type = "prob"`), `vcov()` to obtain the variance-covariance matrix of the parameters of the fitted object, `plot()` to represent graphically the log-likelihood profile over the iterations, and `summary()` to generate the summary of the fitted model. The `AIC()` and the `BIC()` functions are available to obtain the values of the Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), respectively. As for the regressions tests (available in the function `anova()`), we implemented the Wald test to check $H_0 : \delta_{j,k} = 0$ (where $\delta_{j,k}$ is the effect of the $k$th explanatory variable on the $j$th category). In addition, to investigate the significance of terms in the linear predictor, one can obtain the likelihood-ratio test that compares nested models by specifying the two models in the `anova()` function. We also adapted the `step()` function of the **stats** package to incorporate the classical stepwise variable selection for $(r, F, Z)$ models. Similar to the original function, our adaptation employs the AIC criterion in the stepwise algorithm for selecting variables. The resulting function provides forward and backward directions. The user has to define the chosen design for each explanatory variable. Both directions will then conserve these constraints on the coefficients.

Note that the link function $g : \Delta \to \mathbb{R}^{J-1}$ is differentiable if the ratio $\boldsymbol{r} : \Delta \to (0,1)^{J-1}$ and the CDF $F : \mathbb{R} \to (0,1)$ are both differentiable. All the CDFs available in **GLMcat** (see Table 2) are differentiable (i.e., there exists a density function such that $f = F'$). The four ratios are also differentiable, thus, we use the Fisher's scoring algorithm for the estimation of the model. In the following, we present the form of the algorithm in the iteration $t$

$$\boldsymbol{\beta}^{[t+1]} = \boldsymbol{\beta}^{[t]} - \left\{ \mathsf{E}\left( \frac{\partial^2 l}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \right)_{\boldsymbol{\beta}=\boldsymbol{\beta}^{[t]}} \right\}^{-1} \left( \frac{\partial l}{\partial \boldsymbol{\beta}} \right)_{\boldsymbol{\beta}=\boldsymbol{\beta}^{[t]}}.$$

Applying the chain rule to the log-likelihood (given here for one observation) $l = \ln P(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\beta}) = \boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})$ , we obtain the score

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\pi}} \frac{\partial l}{\partial \boldsymbol{\theta}}.$$

Since the response distribution belongs to the exponential family, it becomes

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = Z^\top \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\eta}} \mathsf{COV}(\boldsymbol{Y}|\boldsymbol{x})^{-1}(\boldsymbol{y} - \boldsymbol{\pi}).$$

Then, using the decomposition of the link function presented in Equation 1, we obtain

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = Z^\top \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{r}} \mathsf{COV}(\boldsymbol{Y}|\boldsymbol{x})^{-1}(\boldsymbol{y} - \boldsymbol{\pi}) ,$$

and Fisher's information matrix

$$\mathsf{E}\left( \frac{\partial^2 l}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \right) = -Z^\top \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{r}} \mathsf{COV}(\boldsymbol{Y}|\boldsymbol{x})^{-1} \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{r}^\top} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}^\top} Z.$$

Remark that the Jacobian matrix $\partial \boldsymbol{F}/\partial \boldsymbol{\eta}$ is the diagonal matrix of densities $\{f(\eta_j)\}_{j=1,\ldots,J-1}$; the Jacobian matrices associated to each ratio $\partial \boldsymbol{\pi}/\partial \boldsymbol{r}$ are detailed in Appendix B. Note that the above calculations of the score and the Fisher's information matrices concern only one observation. To obtain the total expected result, the contributions of the $n$ observations have to be added.

Computational difficulties for the maximum likelihood are expected when either complete or quasi complete separation occurs in the dataset, this is due to the fact that the maximum likelihood estimator (MLE) is not unique in that case. Another situation involving such difficulties occurs for the cumulative ratio used together with a complete or partially parallel design; these models are not invertible and the algorithm might fail to converge (more details are given in Section 3). The standard numerical optimization techniques have no way of detecting this problem and will keep iterating until the iteration's bound is reached (Albert and Anderson 1984). The convergence criteria for the Fisher's scoring algorithm is set to be reached in **GLMcat** when

$$\frac{|l(\boldsymbol{\beta}^{[t+1]}) - l(\boldsymbol{\beta}^{[t]})|}{\varepsilon + |l(\boldsymbol{\beta}^{[t+1]})|} > \frac{\varepsilon}{n}, \tag{3}$$

where $\varepsilon = 0.0001$ by default. Thus, the algorithm will stop iterating either when the maximum number of iterations is met, or until the Equation 3 becomes true. In case of convergence problems, an additional strategy is to initialize the model parameters $\boldsymbol{\beta}^{[0]}$ specifying a numerical vector in the argument `control_glmcat`. For the reference, adjacent and sequential ratios, the algorithm is initiated with $\boldsymbol{\beta}^{[0]}$ as the null vector. Conversely, the intercepts of cumulative models are symmetrically and ascendingly defined around 0, thus $\alpha_1^{[0]} < \alpha_2^{[0]} < \ldots < \alpha_{J-1}^{[0]}$. In **GLMcat**, the user can also modify the number of iterations (which by default is 25), and the size of the convergence tolerance given by $\varepsilon$ with the argument `control_glmcat`, for example: `control_glmcat(iterations = 30, epsilon = 0.0001)`. In scenarios where a user aims to employ the Student's distribution but is uncertain about the degrees of freedom, they can set the `find_nu` argument to `TRUE` within the function. This triggers the use of the `optimize()` function, which is a part of the **base** R package. The `optimize()` function combines the golden section search and parabolic interpolation methods (Brent 1973) to locates the minimum or maximum of a single-variable function within a specified interval. In our case, we've defined the search interval as [0.25, 8], and the optimization objective as the log-likelihood.

As Wickham (2015) states, R is a high-level expressive language, and that expressivity comes at a price: speed. In order to improve the speed of the functions in **GLMcat**, we incorporated C++ (Stroustrup 2013) code through the **Rcpp** package (Eddelbuettel and François 2011). The algorithms are implemented in a modular manner, meaning that enhancement or adjustment can be easily extended to all the families of models.

# 3. Models for ordinal responses

Based on the common foundation exposed by the triplet $(r, F, Z)$, it is possible to describe some properties of the models for categorical responses (Peyhardi *et al.* 2015). Such information empowers the practitioner to adequately choose (from a wide range of options) the model that best suits the characteristics of the data. As indicated in the past sections, the link function is composed of $r$ and $F$. By changing either of them, one might obtain improvements in terms of the goodness-of-fit measures. Nevertheless, the performance of a model is not merely measured through the fit. The foremost consideration for choosing a model should be the consistency among the nature of the data, the modeling objectives, and the model's features. In the following, we introduce and illustrate on real datasets, by means of **GLMcat**, the properties of the GLMs for ordinal responses. We intend to guide the practitioner in the selection process of the link function.

## 3.1. Reversibility

To announce the reversibility property of the models for ordinal responses, we need to recall the following definitions introduced by Peyhardi *et al.* (2015):

- The models $(r, F, Z)$ and $(r', F', Z')$ are said to be equivalent if one is a reparameterization of the other, i.e., there exists a bijection $h$ from $\Theta$ to $\Theta'$ such that $r^{-1} \circ \boldsymbol{F}\{Z(\boldsymbol{x})\boldsymbol{\beta}\} = (r')^{-1} \circ \boldsymbol{F}'\{Z'(\boldsymbol{x})h(\boldsymbol{\beta})\}$, for all $\boldsymbol{x} \in \mathcal{X}$, and all $\boldsymbol{\beta} \in \Theta$.
- The models $(r, F, Z)$ and $(r', F', Z')$ are said to be equal if the corresponding distributions of $\boldsymbol{y}|\boldsymbol{x}$ are equal, i.e., if $r^{-1} \circ \boldsymbol{F}\{Z(\boldsymbol{x})\boldsymbol{\beta}\} = (r')^{-1} \circ \boldsymbol{F}'\{Z'(\boldsymbol{x})\boldsymbol{\beta}\}$, for all $\boldsymbol{x} \in \mathcal{X}$, and all $\boldsymbol{\beta} \in \Theta$. Note that the equality between models implies that they are equivalent.
- An $(r, F, Z)$ model is said to be invariant under a permutation $\sigma$ of $\{1, \ldots, J\}$, if it is equivalent to the $(r, F, Z)_\sigma$ model which is defined on the permuted vector $\boldsymbol{\pi}_\sigma = (\pi_{\sigma(1)}, \ldots, \pi_{\sigma(J-1)})$.

On the basis of the above definitions, an $(r, F, Z)$ model is said to be reversible if it is invariant under the reverse permutation $\tilde{\sigma}$ defined by $\tilde{\sigma}(j) = J - j + 1$ for all $j \in \{1, \ldots, J-1\}$. The reversibility property was first studied for cumulative models with some specific distributions by McCullagh (1980). Later, Peyhardi *et al.* (2015) generalized it for all symmetric distributions as well as for the adjacent ratio.

**Proposition 1.** *The (adjacent, F, Z) and the (cumulative, F, Z) models are reversible for all symmetric CDFs F and all the design matrices Z proposed in this package.*

McCullagh (1980) suggests that depending on the application, the reversibility may be seen as an appealing property, for example, when the response is given by an ordered scale.

Moreover, for any CDF $F \in \mathfrak{F}$ and $Z \in \mathfrak{Z}$, we have that:

**Proposition 2.** *The* $(adjacent,\ F,\ Z)_{\tilde{\sigma}}$ *model and the* $(cumulative,\ F,\ Z)_{\tilde{\sigma}}$ *model are respectively equal to the* $(adjacent,\ \tilde{F},\ -\tilde{P}Z)$ *and the* $(cumulative,\ \tilde{F},\ -\tilde{P}Z)$*, where* $\tilde{F}(\eta) = 1 - F(-\eta)$*, and* $\tilde{P}$ *is the restricted reverse permutation matrix of dimension* $J - 1$*:*

$$\tilde{P} = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}. \tag{4}$$

Refer to the Appendix C for the demonstration.

Note that if a CDF is symmetric then $\tilde{F} = F$; for asymmetric distributions, as the Gumbel CDF, $\tilde{F}$ corresponds to its symmetric counterpart, in this example, the Gompertz CDF. For a practical illustration of Proposition 2, consider the observations of the boys' disturbing dreams benchmark dataset presented by Maxwell (1961). This study cross-classified boys by their age $\boldsymbol{x}$ (which corresponds to the mid-point values for each stratum of 2 or 3 years, and it is treated as a continuous explanatory variable), and the severity of their disturbing dreams $Y$ on a four-point scale of increasing severity. The data is available as the object `DisturbedDreams` in the **GLMcat** package. The (*adjacent*, *Gumbel*, *parallel*) model is defined as:

```
R> data("DisturbedDreams", package = "GLMcat")
R> adj_gumbel_p <- glmcat(formula = Level ~ Age, data = DisturbedDreams,
+    ratio = "adjacent", cdf = "gumbel", categories_order = c("Not.severe",
+        "Severe.1", "Severe.2", "Very.severe"))
R> logLik(adj_gumbel_p)

'log Lik.' -279.9612 (df=4)

R> summary(adj_gumbel_p)

Level ~ Age
               ratio    cdf nobs niter    logLik
Model info: adjacent gumbel  223   (7) -279.9612
                        Estimate Std. Error z value Pr(>|z|)
(Intercept) Not.severe   0.22676    0.26157   0.867    0.386
(Intercept) Severe.1    -0.36548    0.24270  -1.506    0.132
(Intercept) Severe.2    -0.33321    0.22899  -1.455    0.146
Age                      0.07146    0.01806   3.957 7.59e-05 ***
```

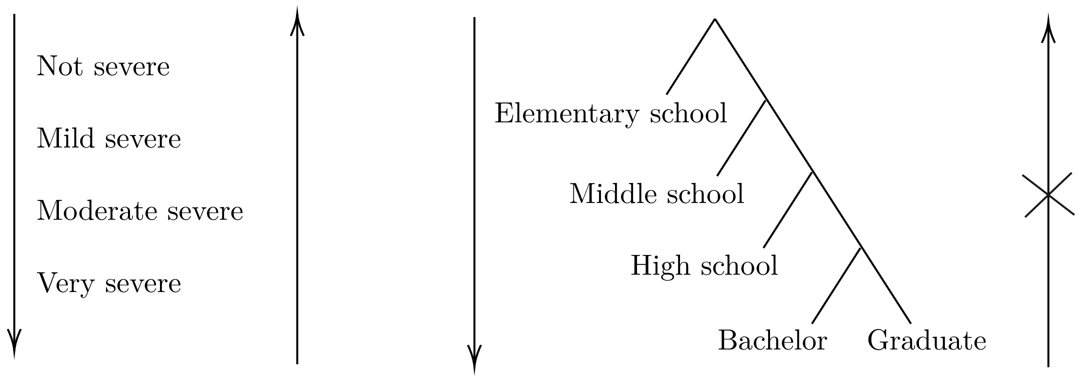


Figure 1: Scale ordering in severity of disturbed dreams versus process ordering in the educational path.

Now, inverting the order of the categories in the argument `categories_order`, and using the symmetric counterpart CDF of the Gumbel, we fit the $(adjacent, Gompertz, parallel)_{\tilde{\sigma}}$ model:

```
R> adj_gompertz_rev <- glmcat(formula = Level ~ Age, data = DisturbedDreams,
+    ratio = "adjacent", cdf = "gompertz", categories_order = c("Very.severe",
+       "Severe.2", "Severe.1", "Not.severe"))
R> logLik(adj_gompertz_rev)

'log Lik.' -279.9612 (df=4)

R> summary(adj_gompertz_rev)

Level ~ Age
               ratio      cdf nobs niter    logLik
Model info: adjacent gompertz  223     7 -279.9612
                         Estimate Std. Error z value Pr(>|z|)
(Intercept) Very.severe  0.33321    0.22899   1.455    0.146
(Intercept) Severe.2     0.36548    0.24270   1.506    0.132
(Intercept) Severe.1    -0.22676    0.26157  -0.867    0.386
Age                     -0.07146    0.01806  -3.957 7.59e-05 ***
```

Note that the estimated parameters of the last model are reversed and with the opposite sign, but the log-likelihood is still the same. This would also be true using any symmetric CDF. Given Proposition 1, the cumulative and the adjacent models are suitable for the type of responses that have an ordering scale associated with their categories. However, the reversibility property is not valid for the sequential models since these are non-invariant under the reverse permutation. The user should consider the sequential ratio if there is a time-related notion or a time-ordered process (which cannot be reversed) implicit in the response. For instance, the education level (see Figure 1) is conditioned on the completion of the previous degrees. The sequential ratio is the only one that takes into account this particularity of the response variable, for this reason, it is commonly employed in survival analysis.

### 3.2. Total invariance

An $(r, F, Z)$ model is said to be totally invariant if it is invariant under all permutations of the response categories. It is well known that the MNL or equivalently, the (*reference*, *logistic*, *complete*) model is totally invariant. Agresti (2010) demonstrated that this model is equivalent to the (*adjacent*, *logistic*, *complete*) model. To illustrate this equivalence, consider the two models: (*reference*, *logistic*, *complete*) and (*adjacent*, *logistic*, *complete*), for the DisturbedDreams dataset:

```
R> mod_ref_log_c <- glmcat(formula = Level ~ Age, ratio = "reference",
+    parallel = FALSE, data = DisturbedDreams, cdf = "logistic")
R> mod_adj_log_c <- glmcat(formula = Level ~ Age, ratio = "adjacent",
+    parallel = FALSE, data = DisturbedDreams, cdf = "logistic")
R> logLik(mod_ref_log_c); logLik(mod_adj_log_c)

'log Lik.' -277.1345 (df=6)
'log Lik.' -277.1345 (df=6)

R> coef(mod_ref_log_c)

(Intercept) Not.severe -2.454
(Intercept) Severe.1   -0.555
(Intercept) Severe.2   -1.125
Age Not.severe          0.310
Age Severe.1            0.060
Age Severe.2            0.112

R> coef(mod_adj_log_c)

(Intercept) Not.severe -1.8998
(Intercept) Severe.1    0.5700
(Intercept) Severe.2   -1.1246
Age Not.severe          0.2500
Age Severe.1           -0.0523
Age Severe.2            0.1123
```

Remark that the log-likelihoods of the last two models are equal but the estimations of the parameters are different. As demonstrated by Peyhardi *et al.* (2015), there exists a matrix $A$ (see Appendix C for details), such that $A\boldsymbol{\alpha} = \boldsymbol{\alpha}'$ for the intercepts, and $A\boldsymbol{\delta} = \boldsymbol{\delta}'$ for the slopes.

```
R> A <- matrix(c(1, 0, 0, -1, 1, 0, 0, -1, 1), nrow = 3)
R> A %*% coef(mod_ref_log_c)[1:3]

[1,] -1.8998
[2,]  0.5700
[3,] -1.1246

R> A %*% coef(mod_ref_log_c)[4:6]
```

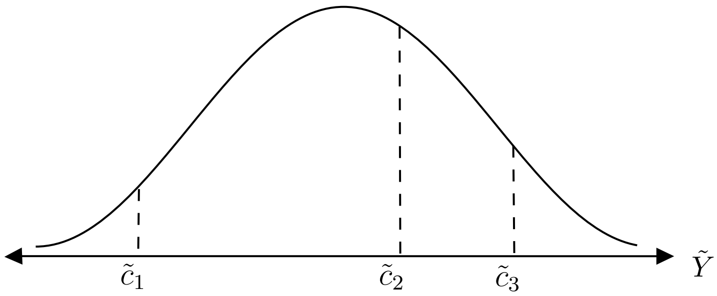


Figure 2: The cumulative model represented through a latent continuous variable.

```
[1,]  0.2500
[2,] -0.0523
[3,]  0.1123
```

As well as the (*reference*, *logistic*, *complete*), the (*adjacent*, *logistic*, *complete*) model is totally invariant and therefore, it is inappropriate for ordinal responses. Apart from this model, any other model in the adjacent family preserves the order assumption, yet, this family has usually been ignored when dealing with ordinal responses.

### 3.3. Latent variable interpretation

As considered by McCullagh (1980) and Bürkner and Vuorre (2019), the (*cumulative*, *logistic*, *proportional*) model can be seen as if the observed $Y$ was originated from the categorization of a latent continuous variable $\tilde{Y}$. This latent variable follows a linear regression model $\tilde{Y} = \tilde{\alpha} + \boldsymbol{x}^\top \tilde{\boldsymbol{\delta}} + \varepsilon$ where $\varepsilon$ is a noise variable with CDF $F$. To model this categorization process, the cumulative ratio assumes that there exists $J-1$ strictly ordered cut-points $-\infty = \tilde{c}_0 < \tilde{c}_1 < \ldots < \tilde{c}_{J-1} < \tilde{c}_J = \infty$ that partition $\tilde{Y}$ into $J$ observable ordered categories of $Y$, i.e.,

$$\{Y = j\} \Leftrightarrow \tilde{c}_{j-1} < \tilde{Y} \leq \tilde{c}_j,$$

for $j = 1, \ldots, J$. The cumulative probabilities are

$$\begin{aligned} \mathsf{P}(Y \leq j | \boldsymbol{x}) &= \mathsf{P}(\tilde{Y} \leq \tilde{c}_j) \\ &= \mathsf{P}(\varepsilon \leq \tilde{c}_j - \tilde{\alpha} - \boldsymbol{x}^\top \tilde{\boldsymbol{\delta}}) \\ &= F(\alpha_j + \boldsymbol{x}^\top \boldsymbol{\delta}) \end{aligned}$$

with $\alpha_j = \tilde{c}_j - \tilde{\alpha}$, and $\boldsymbol{\delta} = -\tilde{\boldsymbol{\delta}}$. We represent this structure (for $J = 4$) in Figure 2, where we can see that

$$\pi_j = \mathsf{P}(\tilde{c}_{j-1} < \tilde{Y} < \tilde{c}_j).$$

The order structure is more easily interpretable using the notion of the latent continuous variable where the categories are considered as successive intervals $(\tilde{c}_{j-1}, \tilde{c}_j]$. Remark that this only holds when the constraint of proportionality is assumed for all explanatory variables. In the other cases (complete or partial parallel design) the interpretation in terms of the latent variable is no longer accurate.

The sequential ratio assumes that the successive choices between category $j$ and the categories over $j$ is determined by the latent variables $\tilde{Y}_j = \tilde{\alpha} + \boldsymbol{x}^\top \tilde{\boldsymbol{\delta}}_j + \varepsilon_j$, for $j = 1, \ldots, J-1$,

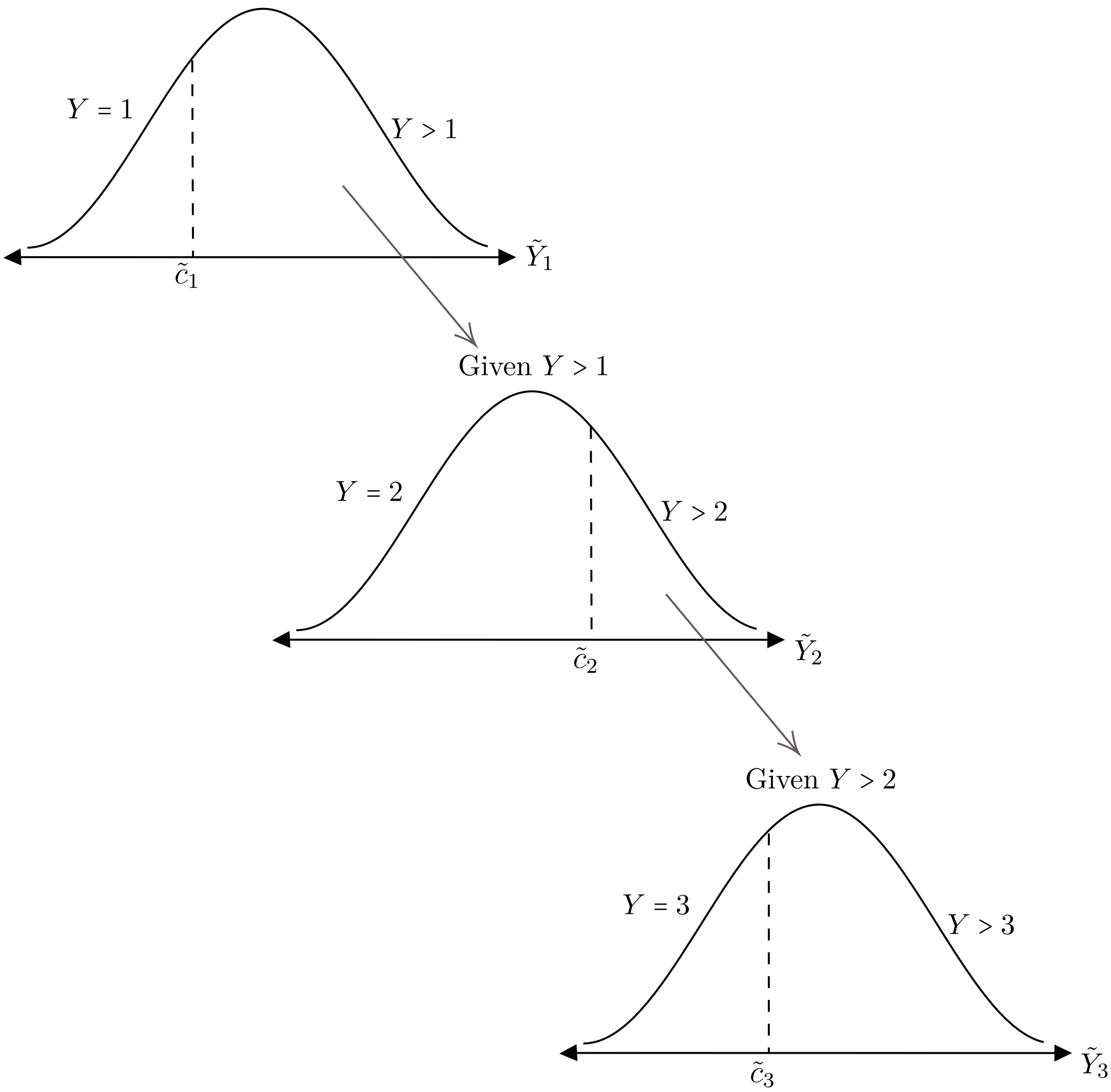


Figure 3: The sequential model represented as latent continuous variables.

where the residuals $\varepsilon_j$ are independent and identically distributed according to the CDF $F$. This sequential mechanism can be viewed as a binary process at each transition, thus, it is appropriate when the assumption of a single underlying latent variable does not hold. We can write then

$$\{Y = j\} = \bigcap_{k=1}^{j-1} \{\tilde{Y}_k > \tilde{c}_k\} \bigcap \{\tilde{Y}_j \leq \tilde{c}_j\},$$

so the conditional probabilities of the event $\{Y = j | Y \geq j\}$ for $j = 1, \ldots, J$ can also be written as $\{Y = j | Y \geq j\} = \{\tilde{Y}_j \leq \tilde{c}_j\}$, then, we have

$$\mathsf{P}(Y = j | Y \geq j; \boldsymbol{x}) := F(\alpha_j + \boldsymbol{x}^\top \boldsymbol{\delta}_j),$$

where $\alpha_j = \tilde{c}_j - \tilde{\alpha}$, and $\boldsymbol{\delta}_j = -\tilde{\boldsymbol{\delta}}_j$. In Figure 3, we illustrated the sequential model with a process that starts from category 1. If $\tilde{Y}_1 \leq \tilde{c}_1$ the process stops and we have $Y = 1$, otherwise, i.e., $\tilde{Y}_1 > \tilde{c}_1$, the process continues and we know that it will at least reach category 2. The

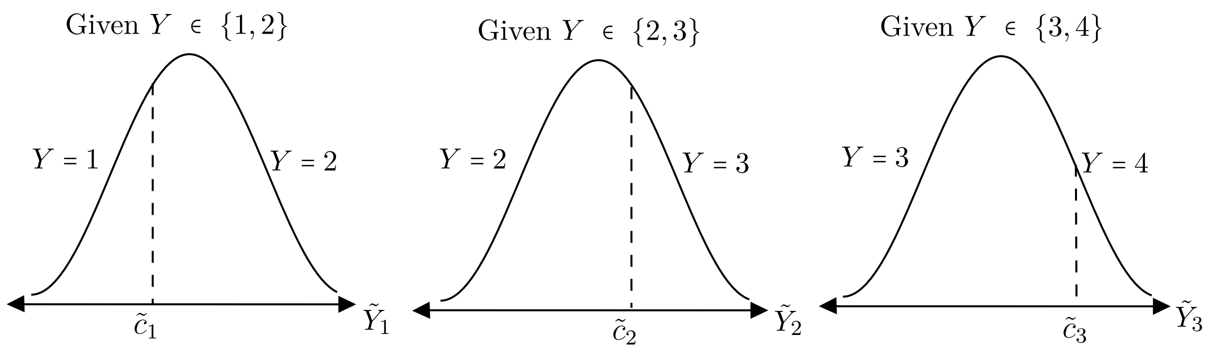


Figure 4: The adjacent model partially represented as a latent continuous variables.

process continues in this way until the last category is reached. In this context, we can represent the probabilities of each category as

$$\pi_j = \mathsf{P}(\tilde{Y}_j < \tilde{c}_j) \prod_{k=1}^{j-1} \mathsf{P}(\tilde{Y}_k > \tilde{c}_k).$$

The transition can be interpreted in terms of the difficulty of reaching the next category. Upper levels can only be achieved if previous levels were visited earlier and not kept. Therefore the model is built around the conditionality principle.

The adjacent ratio describes the probability that category $j$ rather than category $j+1$ is achieved:

$$\{Y = j | Y \in \{j, j+1\}\} = \{\tilde{Y}_j \leq \tilde{c}_j\}$$

for $j = 1, \dots, J-1$. In Figure 4, we represent the adjacent ratio using latent continuous variables. Note that each category $j$ is present in two different latent variables $\tilde{Y}_j$ and $\tilde{Y}_{j+1}$. In contrast to the cumulative and the sequential ratio, $\pi_j$ cannot be written only in terms of the latent variables:

$$\pi_j = \mathsf{P}(\tilde{Y}_j < \tilde{c}_j)(\pi_j + \pi_{j+1}).$$

As a result, this ratio lacks of interpretability since there is not a natural process that leads to its formulation.

## 3.4. Invertibility

An $(r, F, Z)$ model is said to be invertible if its link function is invertible, i.e.,

$$\boldsymbol{\pi} = r^{-1} \circ F(\boldsymbol{\eta}) \in \Delta, \ \forall \boldsymbol{\eta} \in \mathbb{R}^{J-1}.$$

For the cumulative ratio we have

$$\pi_j = F(\eta_j) - F(\eta_{j-1}),$$

and thus, $\eta_{j-1} > \eta_j$ implies $\pi_j < 0$. Therefore, the family of cumulative models is not invertible. To illustrate the case of a non-invertible model, consider the effect on road accident severity caused by the speed limit (`speed_limit`), the road type (`road` and `urban_or_rural_area`), the light (`light_conditions`), the weather conditions (`weather`) of the road where the accident occurred, and the number of casualties (`number_of_casualties`). For this analysis, we used the data from 2019 openly available in `https://data.gov.uk/` and accessible using the **stats19** package (Lovelace, Morgan, Hama, Padgham, Ranzolin, and Sparks 2019). This data set is also available in **GLMcat**. In the presence of the ordered response variable (accident severity with levels: slight, serious, and fatal), the ratio candidates to consider are cumulative, sequential, and adjacent. We first tried to fit the (*cumulative*, *Cauchy*, *complete*) triplet but due to the strong restriction of the cumulative ratio, this attempt failed to converge:

```
R> data("accidents", package = "GLMcat")
R> glmcat(accident_severity ~ road + urban_or_rural_area +
+    day_of_week + number_of_casualties + weather +
+    light_conditions + speed_limit, data = accidents,
+    parallel = FALSE, ratio = "cumulative", cdf = "cauchy")

Warning messages:
1: In .GLMcat(formula = formula, data = data, ratio = ratio, cdf = cdf,  :
  Fisher matrix is not invertible. Check for convergence problems
```

One of the simplest ways of tackling this problem is to impose the constraint $\eta_{j-1} < \eta_j$ through the use of the parallel design. Evidently, the parallel constraint reduces the complexity of the Fisher's scoring algorithm since the condition to preserve the order in the successive iterations is only linked to the intercepts, i.e., $\alpha_{j-1} < \alpha_j$. However, to our knowledge, no previous research has investigated the validity of this constraint in iteration $t$ after having imposed it in iteration 0. A widely used model with the parallel constraint is the odds proportional logit model. Its widespread popularity is due to the fact that it is ideal in terms of interpretation ease and of model parsimony (Abreu, Siqueira, Cardoso, and Caiaffa 2008). However, in practice, the parallel assumption does not usually hold when considering more than one explanatory variable (Lall, Campbell, Walters, Morgan, and MRC CFAS Team 2002), thus, this restrictive assumption is often violated. Continuing with the example, the (*cumulative*, *Cauchy*, *parallel*) model is successfully fitted through:

```
R> cum_cau <- glmcat(accident_severity ~ road + urban_or_rural_area +
+    day_of_week + number_of_casualties + weather +
+    light_conditions + speed_limit, data = accidents,
+    parallel = TRUE, ratio = "cumulative", cdf = "cauchy")
R> summary(cum_cau)

accident_severity ~ road + urban_or_rural_area + day_of_week +
    number_of_casualties + weather + light_conditions + speed_limit
                  ratio    cdf   nobs niter logLik
Model info: cumulative cauchy 109577  (10) -62593
                              Estimate Std. Error z value Pr(>|z|)
(Intercept) Slight            1.819854   0.103695   17.55  < 2e-16 ***
(Intercept) Serious          22.134261   0.544747   40.63  < 2e-16 ***
roadOne way street           -0.082418   0.096072   -0.86  0.39096
roadRoundabout                0.247637   0.066977    3.70  0.00022 ***
roadSingle carriageway       -0.502254   0.030708  -16.36  < 2e-16 ***
roadSlip road                 0.431389   0.115313    3.74  0.00018 ***
```

```
urban_or_rural_areaUrban       0.265489  0.026713    9.94  < 2e-16 ***
day_of_weekMonday              0.033254  0.035015    0.95  0.34226
day_of_weekSaturday           -0.096729  0.033531   -2.88  0.00392 **
day_of_weekSunday             -0.196414  0.033577   -5.85  4.9e-09 ***
day_of_weekThursday           -0.036032  0.033538   -1.07  0.28266
day_of_weekTuesday             0.045625  0.034784    1.31  0.18963
day_of_weekWednesday          -0.007348  0.034093   -0.22  0.82935
number_of_casualties          -0.155693  0.009172  -16.98  < 2e-16 ***
weatherFine no high winds      0.106271  0.077432    1.37  0.16993
weatherFog or mist             0.370257  0.169997    2.18  0.02940 *
weatherRaining + high winds    0.081563  0.105066    0.78  0.43757
weatherRaining no high winds   0.226007  0.081171    2.78  0.00536 **
weatherSnowing                 0.549765  0.207627    2.65  0.00810 **
light_conditionsDaylight       0.209250  0.019735   10.60  < 2e-16 ***
speed_limit                   -0.010764  0.000921  -11.69  < 2e-16 ***
```

On the other hand, we observe that the adjacent and the sequential models are both invertible using any form of the linear predictor. We can write the probabilities of the $(adjacent, F, Z)$ models as:

$$\pi_j = \frac{\prod_{k=j}^{J-1} F(\eta_k)/(1-F(\eta_k))}{1+\sum_{k=1}^{J-1} F(\eta_k)/(1-F(\eta_k))} \, ,$$

and the probabilities of $(sequential, F, Z)$ models in the form:

$$\pi_j = F(\eta_j)\prod_{k=1}^{j-1}(1-F(\eta_k)).$$

In both cases, one can readily identify that $0 < \pi_j < 1$ for all $j \in \{1, \dots, J-1\}$ such that $\sum_{j=1}^{J} \pi_j = 1$. If the slope effect is expected to be different for each category and the cumulative ratio fails to fit the model, the practitioner should consider the adjacent or sequential ratios instead. For our example, since the order of the response categories is not time-dependent, we fit the (*adjacent*, *Cauchy*, *complete*) model obtaining:

```
R> adj_cau <- glmcat(accident_severity ~ road + urban_or_rural_area +
+    day_of_week + number_of_casualties + weather +
+    light_conditions + speed_limit, data = accidents,
+    parallel = FALSE, ratio = "adjacent", cdf = "cauchy")
R> summary(adj_cau)

accident_severity ~ road + urban_or_rural_area + day_of_week +
    number_of_casualties + weather + light_conditions + speed_limit
                ratio    cdf   nobs niter logLik
Model info: adjacent cauchy 109577  (14) -62060
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept) Slight                   1.97092    0.11264   17.50  < 2e-16 ***
(Intercept) Serious                  9.89386    0.95812   10.33  < 2e-16 ***
roadOne way street Slight           -0.09133    0.10924   -0.84  0.40312
roadOne way street Serious           3.46606    3.39346    1.02  0.30707
```

```
roadRoundabout Slight                0.27694   0.07647    3.62  0.00029 ***
roadRoundabout Serious               6.13180   2.42201    2.53  0.01135 *
roadSingle carriageway Slight       -0.55281   0.03403  -16.25  < 2e-16 ***
roadSingle carriageway Serious      -1.06469   0.22537   -4.72  2.3e-06 ***
roadSlip road Slight                 0.48098   0.13217    3.64  0.00027 ***
roadSlip road Serious               -0.36728   0.47703   -0.77  0.44134
urban_or_rural_areaUrban Slight      0.29358   0.02930   10.02  < 2e-16 ***
urban_or_rural_areaUrban Serious     0.58181   0.35379    1.64  0.10007
day_of_weekMonday Slight             0.03669   0.03844    0.95  0.33982
day_of_weekMonday Serious           -0.37198   0.31697   -1.17  0.24058
day_of_weekSaturday Slight          -0.10704   0.03657   -2.93  0.00342 **
day_of_weekSaturday Serious         -0.51803   0.29851   -1.74  0.08267 .
day_of_weekSunday Slight            -0.21198   0.03652   -5.80  6.4e-09 ***
day_of_weekSunday Serious           -0.44009   0.30439   -1.45  0.14822
day_of_weekThursday Slight          -0.03646   0.03678   -0.99  0.32151
day_of_weekThursday Serious         -0.20658   0.32653   -0.63  0.52695
day_of_weekTuesday Slight            0.05491   0.03830    1.43  0.15164
day_of_weekTuesday Serious          -0.50526   0.30869   -1.64  0.10168
day_of_weekWednesday Slight         -0.00162   0.03749   -0.04  0.96559
day_of_weekWednesday Serious        -0.58575   0.30017   -1.95  0.05101 .
number_of_casualties Slight         -0.17765   0.00967  -18.37  < 2e-16 ***
number_of_casualties Serious        -0.22305   0.03731   -5.98  2.3e-09 ***
weatherFine no high winds Slight     0.12416   0.08344    1.49  0.13675
weatherFine no high winds Serious    0.60544   0.42456    1.43  0.15385
weatherFog or mist Slight            0.38626   0.18480    2.09  0.03660 *
weatherFog or mist Serious          -0.43593   0.54573   -0.80  0.42441
weatherRaining + high winds Slight   0.10872   0.11401    0.95  0.34029
weatherRaining + high winds Serious  0.28617   0.56831    0.50  0.61458
weatherRaining no high winds Slight  0.25111   0.08761    2.87  0.00415 **
weatherRaining no high winds Serious 1.73089   0.53283    3.25  0.00116 **
weatherSnowing Slight                0.59895   0.22914    2.61  0.00895 **
weatherSnowing Serious               2.17619   2.17244    1.00  0.31648
light_conditionsDaylight Slight      0.23617   0.02150   10.99  < 2e-16 ***
light_conditionsDaylight Serious     0.86121   0.15687    5.49  4.0e-08 ***
speed_limit Slight                  -0.01167   0.00101  -11.60  < 2e-16 ***
speed_limit Serious                 -0.11368   0.01157   -9.82  < 2e-16 ***
```

In the analysis of the road accidents, the complete design could not be fitted using the cumulative model. For this reason, we considered the (*cumulative*, *Cauchy*, *parallel*) model for which the AIC results to be 125227. Then, we used the adjacent ratio aiming to investigate the complete design. From the reported output of the (*adjacent*, *Cauchy*, *complete*) model and based on the p-values, one can observe that many of the explanatory variables are significant with the complete design, thus for each category. Although the number of model parameters was increased, we observed an improvement in terms of the AIC which for the adjacent model was 124199.

| | Reversibility | Latent variable interpretation | Invertibility | Not totally invariant |
|---|---|---|---|---|
| Sequential | | ✓ | ✓ | ✓ |
| Cumulative | ✓ | ✓(i) | | ✓ |
| Adjacent | ✓ | | ✓ | ✓(ii) |

Table 4: Properties of the ratios for ordinal responses; shaded cells indicate that the property is valid. (i) true only with the parallel design $Z_p$. (ii) true only if $F$ is different from *Logistic* or if $Z \neq Z_p$.

### 3.5. Choice of an ordinal model

Several authors have suggested that the choice of the model to fit ordinal responses should correspond to the underlying nature of the response variable (Ananth and Kleinbaum 1997; O'Connell 2006; Agresti 2010). On the basis of the above-mentioned properties, we can define some general guidelines for this choice. Firstly, it is important to differentiate whether the ordinal variable has a temporal foundation associated with the occurrence of the categories (time-ordered process); or if it was drawn from an ordered scale which would still be interpretable, even if the order of the categories is reversed (Figure 1 illustrates an example of this differentiation). In the time-ordered process scenario, it is assumed that to reach category $j$ it was necessary to have visited the previous categories $1, \ldots, j-1$. Consider the example of the level of education attained by different people. Following a traditional academic path, it is possible to attend high school only after the completion of both elementary and middle school. In this case, the sequential ratio would be the best option to work with, since it is the one that best captures this dynamic process.

For the ordered scale response variables, either the cumulative or adjacent ratio can be used since they are reversible. However, the adjacent ratio is invertible but there is no interpretation in terms of a latent variable. By contrast, the cumulative ratio relies on the latent variable formulation, but, it is not invertible (see Table 4). In practice, this means that when the practitioner wants to specify either a complete or a partial parallel design, some computation problems may occur when using the cumulative ratio. Moreover, the interpretation via a latent variable does not hold for the cumulative ratio with a design different from the parallel. Therefore, as the adjacent ratio is invertible, the adjacent family of models should be preferred. Still, the cumulative models are the most widely used in the literature. Perhaps, the unpopularity of the adjacent ratio is due to the fact that only the logistic CDF is proposed in most softwares. Moreover, note that the (*adjacent*, *logistic*, *complete*) model is not appropriate for ordered responses due to its total invariance property.

As for the matrix design, multiple alternatives can be considered. Some researchers prefer to start by using the parallel design for all the explanatory variables. If the model fits poorly they might include separate effects by considering the complete design for some or all of the explanatory variables. Other options to address the concern of an inadequate parallel assumption are using different CDFs or adding additional terms to the linear predictor. Furthermore, models with different designs can be compared using the AIC and/or the BIC as measures for parsimony. A conventional technique that aims to minimize some of these criteria is the stepwise variable selection, also available in **GLMcat**. Likewise, several options

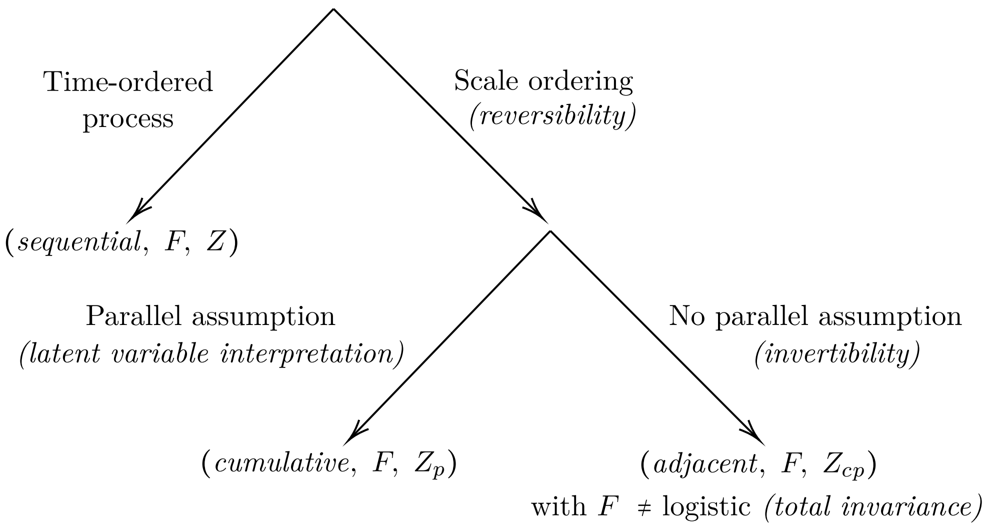


Figure 5: Schematic guide for choosing the appropriate ratio according to the characteristics of the response.

can be used for the CDF component. As mentioned above, the comparison of parameter estimates requires special attention if different distributions are specified. Furthermore, the assumptions of the model on the response are strongly shaped by the choice of the CDF. This is the case of the interaction between the adjacent ratio and the logistic CDF.

Figure 5 compiles the key points we have presented in this section. In summary, we recommend to use a sequential model if there is a time-ordered process among categories. If there is a scale ordering, use a cumulative or adjacent model since they are both reversible. Given that the interpretability through latent variables can be advantageous, we suggest favoring the use of a cumulative model whenever the assumption of parallelism is valid. Otherwise, opt for an adjacent model since it is invertible, and we urge to not use the logistic CDF to avoid the total invariance. We intended to give some general recommendations, however, each analysis has its own particularities which should be addressed from each of the three angles specified by the ratio $r$, the CDF $F$, and the design matrix $Z$. We encourage the user to fit and compare a set of models under different criteria in order to find the $(r, F, Z)$ triplet that best approaches their research questions.

## 4. Models for nominal responses

The MNL is the most popular regression model for categorical responses. In the case of a nominal response, it is often the only model available; except in discrete choice (DC) theory where some extensions have been proposed. In this specific DC framework, the MNL can be interpreted in terms of an underlying behavioral model, the so-called random utility maximization (RUM) model, i.e., $\mathsf{P}(Y = j) = \mathsf{P}(U_j = \max_k U_k)$, where $U_j = \eta_j + \varepsilon_j$ and $\varepsilon_j$'s are independently Gumbel distributed. The $U_j$ associated with each alternative $j$ (category $j$) is called random utility. Two classical extensions are frequently used as RUM models: the multinomial probit (MNP) model, for which $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_J)$ follows a multivariate normal distribution, and the nested logit (NL) model, for which the residuals $\varepsilon_j$ are independent and follow a generalized Gumbel distribution. There are some difficulties with the interpretation and the inference of these models. Since the MLE of the NL model cannot be directly obtained, the model estimation is computed either simultaneously (best alternative when there are less than four nested levels) or sequentially (which might lead to a suboptimal log-likelihood at convergence); more details about this are given by Forinash and Koppelman (1993) and Louviere, Hensher, Swait, and Adamowicz (2000). On the other hand, the estimation of MNP models can be complex (specially when $J > 3$) due to the underlying multidimensional integrals of the multivariate normal density function (Geweke, Keane, and Runkle 1994).

We propose to use an extension of the MNL in the DC framework based on the reference model's family: ($reference$, $F$, $Z$), where $Z$ can take into account variables specific to the alternatives $\{\boldsymbol{\omega}_j\}_{j=1,\ldots,J}$. The linear predictor takes the general form: $\eta_j = \alpha_j + \boldsymbol{x}^\top \boldsymbol{\delta}_j + (\boldsymbol{\omega}_j - \boldsymbol{\omega}_J)^\top \boldsymbol{\gamma}$, for $j = 1, \ldots, J-1$, thus, the design matrix, where $\tilde{\boldsymbol{\omega}}_j = \boldsymbol{\omega}_j - \boldsymbol{\omega}_J$, has the form:

$$\begin{pmatrix} 1 & & \boldsymbol{x}^\top & & \tilde{\boldsymbol{\omega}}_1^\top \\ & \ddots & & \ddots & \vdots \\ & & 1 & & \boldsymbol{x}^\top & \tilde{\boldsymbol{\omega}}_{J-1}^\top \end{pmatrix}.$$

Note that these models are invariant only under the permutation that fixes the reference alternative (see Peyhardi *et al.* 2015, for details). In other words, contrary to the MNL, changing the reference alternative leads to a different model (except if $F$ = logistic). The advantages of the reference models (versus MNP or NL) are:

- they include MNL as a special case ($F$ = logistic),
- their simple inference procedure (Fisher's scoring algorithm),
- their simple interpretation since each alternative is compared to a reference alternative $\frac{\pi_j}{\pi_j + \pi_J} = F(\eta_j)$.

Another good property is the invertibility which is evident from writing the probabilities of the model in the form:

$$\pi_j = \frac{F(\eta_j)/(1 - F(\eta_j))}{1 + \sum_{k=1}^{J-1} F(\eta_k)/(1 - F(\eta_k))}.$$

It should be remarked that the reference models are DC models but not RUM models. Moreover, the ($reference$, $normal$, $Z$) model is different from the MNP model.

We propose to use the family of ($reference$, $Student(\nu)$, $Z$) models, which is an alternative that grants robustness and flexibility through the Student CDF. Indeed, Peyhardi (2020) showed that the influence function is bounded with the Student CDF (contrary to the logistic or normal CDFs). Consequently, these models are less sensitive to outliers than the MNL, in addition, they seem to be less sensitive to noisy explanatory variables. The Student CDF itself generates a family of models as different fits are expected when changing $\nu$. The flexibility we previously mentioned, lies in the increase of the range of possible CDFs to consider as part of the link function. For instance, the three most popular link functions can be obtained with

- $\nu = 1 \Rightarrow F_\nu$ = Cauchy,
- $\nu = 8 \Rightarrow F_\nu \simeq$ logistic, and
- $\nu \to \infty \Rightarrow F_\nu$ = normal.

The Student distribution has been further extended by using a non-centrality parameter $\mu$. This generalization is known as the non-central $t$ distribution. The resulting CDF is also available in **GLMcat** and can be used by specifying its parameters: `cdf = list("noncentralt", df = 5, mu = 2)`. Note that the non-central $t$ distribution is asymmetric unless $\mu = 0$ (in which case it is equivalent to the Student CDF). A detailed description of this distribution can be found in Johnson, Kotz, and Balakrishnan (1995) and its PDF is recalled in Appendix A.

To estimate these models in **GLMcat**, we create the function `discrete_cm()` which requires data in a long format (an example is given in the following). Thus, for each individual (or decision-maker), there are multiple observations (rows), one for each of the alternatives the individual could choose. We call the group of observations for an individual a case. Each case represents a single statistical observation (although it comprises multiple observations), and the identification column of the $n$ cases should be specified in the argument `case_id`. The user must be aware that the `discrete_cm()` function has been built for the particular case of explanatory variables specific to the alternatives. If not required, the user can call the `glmcat()` function using the reference ratio.

## 4.1. Application

Consider the dataset studied by Louviere *et al.* (2000) in which 210 passengers choose one travel mode among the $J = 4$ options: air, train, bus, and car (available in **GLMcat** as the `TravelChoice` object). In this analysis, the individual's attributes are the household income (hinc) and the traveling group size (psize). The alternative specific attributes for each travel mode are the generalized cost (gc) and the terminal waiting time (ttme). The dataset has a long format, i.e., the variables concerning the $n$ individuals are detailed in $n \times J$ lines; an example for the first two individuals is:

```
R> head(TravelChoice, 8)
```

```
  indv  mode choice ttme invc invt gc hinc psize
1    1   air  FALSE   69   59  100 70   35     1
2    1 train  FALSE   34   31  372 71   35     1
3    1   bus  FALSE   35   25  417 70   35     1
4    1   car   TRUE    0   10  180 30   35     1
5    2   air  FALSE   64   58   68 68   30     2
6    2 train  FALSE   44   31  354 84   30     2
7    2   bus  FALSE   53   25  399 85   30     2
8    2   car   TRUE    0   11  255 50   30     2
```

In the following, we estimate and compare a set of models with different CDFs and with different specifications of the reference category.

### *Logistic CDF*

We first estimate the (*reference*, *logistic*, $Z^{(1)}_{car}$) model (which corresponds to the MNL) considering `car` as the reference category, the associated design matrix where $h$ represents `hinc`,

$c$ for gc, $t$ for ttme, and $p$ for psize, is:

$$Z_{car}^{(1)} = \begin{pmatrix} 1 & 0 & 0 & h & 0 & 0 & p & 0 & 0 & c_{air} - c_{car} & t_{air} - t_{car} \\ 0 & 1 & 0 & 0 & h & 0 & 0 & p & 0 & c_{bus} - c_{car} & t_{bus} - t_{car} \\ 0 & 0 & 1 & 0 & 0 & h & 0 & 0 & p & c_{train} - c_{car} & t_{train} - t_{car} \end{pmatrix}.$$

```
R> logistic_car <- discrete_cm(formula = choice ~ hinc +
+    psize + gc + ttme, case_id = "indv", alternatives = "mode",
+    reference = "car", data = TravelChoice, alternative_specific = c("gc",
+       "ttme"), cdf = "logistic")
R> summary(logistic_car)
```

```
 "choice ~ hinc + psize + gc + ttme + indv + mode"
                ratio      cdf nobs niter    logLik
Model info: reference logistic  210   (5) -177.4541
                    Estimate Std. Error z value Pr(>|z|)
X.Intercept. air    7.873608   0.986848   7.979 1.48e-15 ***
X.Intercept. bus    4.433192   0.778334   5.696 1.23e-08 ***
X.Intercept. train  5.559205   0.699139   7.952 1.84e-15 ***
hinc air            0.004071   0.012725   0.320 0.749020
hinc bus           -0.023324   0.016297  -1.431 0.152391
hinc train         -0.055185   0.014482  -3.810 0.000139 ***
psize air          -1.027423   0.265657  -3.867 0.000110 ***
psize bus          -0.030010   0.333977  -0.090 0.928402
psize train         0.302395   0.225616   1.340 0.180144
gc                 -0.019685   0.005401  -3.644 0.000268 ***
ttme               -0.101566   0.011231  -9.044  < 2e-16 ***
```

```
R> logLik(logistic_car)
```

```
'log Lik.' -177.4541 (df=11)
```

A more specific design was studied by Louviere *et al.* (2000, p. 157) and Greene (2003, p. 730). These analyses set the effect of the variables hinc and psize exclusively for the category air, i.e.,

$$\eta_j = \alpha_j + \boldsymbol{x}^\top \boldsymbol{\delta}_{air} \mathbb{1}_{j=air} + (\boldsymbol{\omega}_j - \boldsymbol{\omega}_{car})^\top \boldsymbol{\gamma} \tag{5}$$

for $j \in \{\text{air, bus, train}\}$. Hence, the associated design matrix is:

$$\begin{pmatrix} 1 & 0 & 0 & h & p & c_{air} - c_{car} & t_{air} - t_{car} \\ 0 & 1 & 0 & 0 & 0 & c_{bus} - c_{car} & t_{bus} - t_{car} \\ 0 & 0 & 1 & 0 & 0 & c_{train} - c_{car} & t_{train} - t_{car} \end{pmatrix}.$$

As far as we know, there is no other package in R to fit this particular design. In **GLMcat**, we can fit this model with the lines of code:

```
R> logistic_car_alt <- discrete_cm(formula = choice ~
+    hinc[air] + psize[air] + gc + ttme, case_id = "indv",
+    alternatives = "mode", reference = "car", data = TravelChoice,
+    alternative_specific = c("gc", "ttme"), cdf = "logistic")
R> summary(logistic_car_alt)
```

```
"choice ~ hinc[air] + psize[air] + gc + ttme + indv + mode"
                ratio      cdf nobs niter    logLik
Model info: reference logistic  210   (5) -185.9149
                    Estimate Std. Error z value Pr(>|z|)
X.Intercept. air    7.334807   0.946436   7.750 9.19e-15 ***
X.Intercept. bus    3.591702   0.475771   7.549 4.38e-14 ***
X.Intercept. train  4.371913   0.478124   9.144  < 2e-16 ***
hinc air            0.023815   0.011189   2.128   0.0333 *
psize air          -1.173817   0.258133  -4.547 5.43e-06 ***
gc                 -0.023507   0.005084  -4.624 3.76e-06 ***
ttme               -0.100213   0.010543  -9.505  < 2e-16 ***
```

```
R> logLik(logistic_car_alt)
```

```
'log Lik.' -185.9149 (df=7)
```

By applying the AIC-based stepwise algorithm to the model with the logistic CDF, all four explanatory variables were found to have discriminative properties. This highlights the suitability of the full model with an AIC value of 385.83. Refer to the detailed experimentation in the package's vignette: *An example of variable selection in* ***GLMcat***.

*Student CDF*

We now employ the Student CDF. The design specified by the linear predictor in Equation 5 depends on the reference alternative $j_0$. Since reference models are sensitive to changes in the reference alternative, it is necessary to select $j_0$ appropriately.

To this end, we set the parameter `find_nu = TRUE` in the `discrete_cm()` function, as $\nu$ is not known a priori and can be determined using the method outlined in Section 2.5. Through this approach, we found that the highest log-likelihood $-146.7092$ was achieved when using `car` as the reference alternative, with a degree of freedom of $\nu = 0.490052$.

```
R> mod_car <- discrete_cm(formula = choice ~ hinc +
+    psize + gc + ttme, case_id = "indv", alternatives = "mode",
+    reference = "car", alternative_specific = c("gc", "ttme"),
+    normalization = 0.95, data = TravelChoice, cdf = "student",
+    find_nu = TRUE)
R> mod_car$cdf
```

```
$cdf
[1] "student"

$freedom_degrees
[1] 0.490052

$mu
[1] 0
```

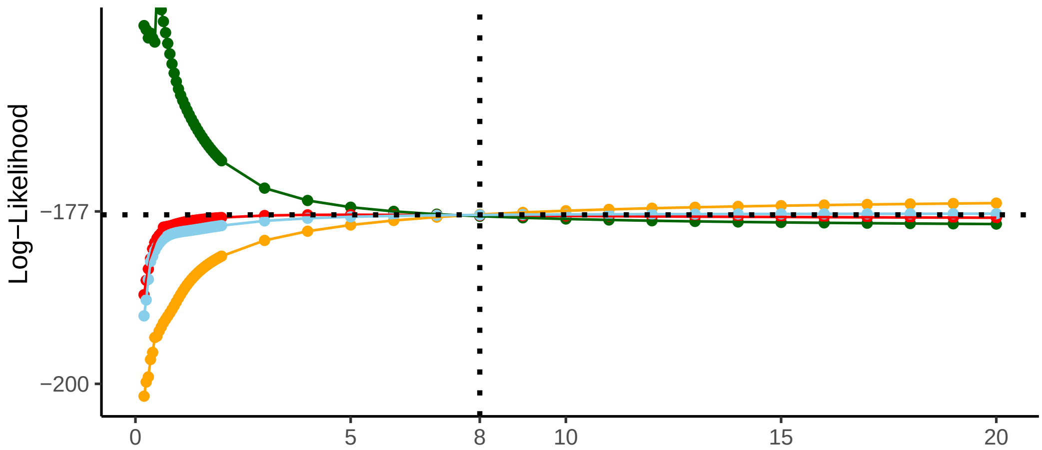


Figure 6: Log-likelihood curves for models with reference category: car (green), bus (yellow), train (red) and air (blue), and with $\nu$ with a 0.05-step from 0.2 to 2 and an integer-step from 2 to 20.

```
R> summary(mod_car, normalized = TRUE)
```

```
Normalized coefficients with s0 =  0.06599551
[1] "choice ~ hinc + psize + gc + ttme + indv + mode"
                ratio      cdf nobs niter    logLik
Model info: reference student  210    12 -146.7092
                    Estimate Std. Error z value Pr(>|z|)
X.Intercept. air    5.422391   1.944758   2.788  0.00530 **
X.Intercept. bus    2.495377   0.872363   2.860  0.00423 **
X.Intercept. train  2.737738   0.947959   2.888  0.00388 **
hinc air            0.007033   0.007077   0.994  0.32035
hinc bus           -0.002872   0.010188  -0.282  0.77805
hinc train         -0.006442   0.007005  -0.920  0.35773
psize air          -0.238991   0.222474  -1.074  0.28271
psize bus           0.364196   0.331718   1.098  0.27224
psize train         0.265797   0.128758   2.064  0.03899 *
gc                 -0.004722   0.001835  -2.574  0.01006 *
ttme               -0.087668   0.029787  -2.943  0.00325 **
```

Additionally, we plotted the curves to visualize how the log-likelihood varies with changes in $\nu$ across different alternatives. As shown in Figure 6, when $\nu = 8$, the log-likelihoods for the four models (one for each alternative) converge to similar values around −177. This result is expected, as the logistic CDF, which approximates the fit of Student(8), is the only model that offers invariance under all permutations of alternatives.

Based on the previous results and a significance level of 0.01, only terminal waiting time significantly impacts the choice of travel mode. Notably, the model with this single explanatory variable has a log-likelihood of −148.7015, which is close to the log-likelihood of −146.709 for the model with all other explanatory variables.

```
R> dis_stu_2 <- discrete_cm(formula = choice ~ ttme, case_id = "indv",
+    alternatives = "mode", reference = "car", data = TravelChoice,
+    alternative_specific = "ttme", cdf = list("student",
+       df = 0.490052))
R> logLik(dis_stu_2)

 'log Lik.' -148.7015 (df=4)
```

Note that not all Wald tests exhibit significance, unlike the case with the logistic CDF. It is may be due to a numerical problem since it is known that the $CDF$ and especially the PDF evaluation is unstable when the $\nu$ is near to zero. Indeed Fisher's information matrix could be poorly estimated and thus the Wald tests could not reflect the reality. To determine the discriminative nature of explanatory variables, a variable selection algorithm based on AIC is employed, encompassing both backward and forward approaches. Strikingly, both algorithms converge to the same model, with only the `ttme` variable serving as the explanatory factor. A direct comparison of AIC values between this Student model and the conventional multinomial logit (MNL) model provides a conclusive insight: solely the `ttme` variable holds discriminative significance in the context of travel mode choice.

*Back to the dataset*

This raises the question of whether the choice of transport mode can be completely determined by the ttme variable.

Let see the raw dataset, with only the choice (first column) and the `ttme` variable (four columns respectively for air, train, bus and car). Here one row correponds to one observation and all the rows have been ordered according to the `ttme` values in order to ease the reading.

```
R> wide_data <- TravelChoice %>%
+    select(individual = indv, mode, travel_time = ttme) %>%
+    spread(key = mode, value = travel_time)
R> sorted_data <- joined_data %>%
+    arrange(desc(air), desc(train), desc(bus), desc(car))
R> head(sorted_data[,-1], 20)

   air bus car train  mode
1   99  35   0    34   air
2   90  53   0    44   air
3   90  53   0    44   air
4   90  53   0    44   air
5   90  35   0    34   air
6   90  35   0    34   air
7   85  53   0    44   air
8   80  35   0    34   air
9   75  53   0    44   air
10  75  35   0    34   air
11  75  35   0    34   air
12  75  35   0    34   air
```

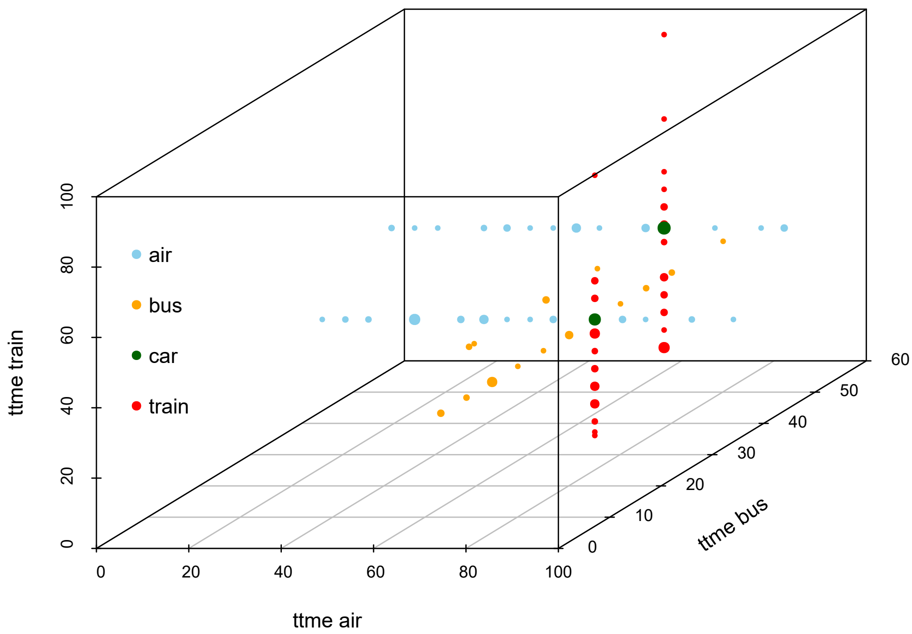


Figure 7: Three-dimensional representation of observed terminal time values. The sizes of the points are proportional to the number of individuals who chose the travel option among: air (blue), bus (yellow), car (green), train (red).

```
13   69   35    0     75  train
14   69   35    0     45  train
15   69   35    0     45  train
16   69   35    0     45  train
17   69   35    0     40  train
18   69   35    0     40  train
19   69   35    0     40  train
20   69   60    0     34    bus
```

Remark that if two observations (two rows) share identical terminal waiting times across all modes of transportation (air, bus, car, and train) then the travel mode choice is the same. This pattern is not limited to the 20 cases displayed; it extends to the entire dataset of 210 observations. As demonstrated in the following code snippet, a single mode of transport is consistently chosen based on the distinct `ttme` combinations for air, bus, car, and train:

```
R> mode_summary <- sorted_data %>%
+    group_by(air, train, bus, car) %>%
+    summarise(modes_count = n_distinct(mode)) %>%
+    ungroup
R> summary(mode_summary$modes_count)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      1       1       1       1       1       1 
```

It means that, the `ttme` column alone can conclusively predict the selected mode of travel, establishing a completely deterministic relationship. If the reader is not convinced, let's visualize the dataset in three dimensions. Since the terminal waiting time for the alternative car is null, it is possible to represent the different values of this variable by points in three dimensions (air, bus, and train) with a color indicating the observed travel mode.

The reader can visualize in Figure 7 that the `ttme` values are completely artificial. Note that there are only two triplets ($\texttt{ttme}_{air}$, $\texttt{ttme}_{bus}$, $\texttt{ttme}_{train}$) for which users choose car, these are: $(69, 35, 34)$ and $(64, 53, 44)$. These points are the intersection of the lines formed by the other observations. Random observations should not be align like this, they should be more chaotic. Even if these lines of `ttme` values corresponds to an experiment and not to real observations, they should be mixed by different colors. Concretely it is impossible to observe such a dataset.

In contrast to the logistic CDF, the Student CDF allows us to discover the completely artificial nature of this classical dataset. This was possible because the Student CDF seems to be more robust to outliers and noise variables (for more details see Peyhardi (2020)).

# 5. Discussion

Liu and Agresti (2005) presented an overview of developments in the analysis of ordinal responses. In their final comments, they highlighted that the current main challenge is to make these methods better known to researchers who commonly encounter this kind of data. Up to now, the models for categorical responses have been popularized in different disciplines separately. We consider that once all the models are assembled, their specific characteristics can be better understood and, thus, users can readily compare and choose a solution tailored to the objectives of their analysis. In the present article, we illustrated a generalized modeling framework for categorical responses, while introducing an R package that encompasses all these models. The contributions presented in this paper have wide applicability given that several fields of research and industry deal extensively with categorical responses. We discussed the properties of the different families of models, as well as the relevance of the choice of both the CDF and the linear predictor's form. With the **GLMcat** package, it is now computationally possible to test a variety of categorical regression models using one single function. We consider that this tool allows to popularize the area of categorical data regression which has not been yet widespread on a large scale through non-logistic models.

Although the most popular CDFs often result in "similar" fits, this does not imply that all CDFs are essentially equivalent when fixing the ratio. In distributions such as the Pregibon (based on the generalized Tukey family) or the non-central *t*, some parameters control the symmetry, the heaviness of the tails, and/or the skewness of the distribution. Hence, one extension would be to consider an algorithm to estimate such parameters. The vast set of new possible CDFs enlarges the toolkit for modeling categorical responses, and with their use, subtle details might be uncovered as illustrated in the example of Section 4. An advantage of the modularized architecture of the package is that it facilitates the inclusion of additional CDFs which will be immediately available for all four model families.

The hierarchical structure of nominal, ordinal, or partially ordinal responses has been already studied among others by Zhang and Ip (2012) and Peyhardi, Trottier, and Guédon (2016). Based on the presented methodology, we can consider the $(r, F, Z)$ triplets as basic units of a hierarchically structured model. This general and flexible model allows taking into account possible relations among response categories. The hierarchical model is then defined by a partition tree where, for each non-terminal node, an $(r, F, Z)$ model is specified. Remark that in this case, the link function would be composed of the tree partition, the set of ratios, and the CDFs specified for the non-terminal nodes.

The GLM presented in Section 2 can be extended to include random effects. Some authors have already made this extension for particular models in the context of categorical responses (see Hartzel, Agresti, and Caffo 2001; Coull and Agresti 2000; Tutz and Hennevogl 1996). The implementation of generalized linear mixed models is envisaged for the $(r, F, Z)$ in **GLMcat**. The **mgcv** (Wood 2011) package provides a unified framework for modeling categorical responses and allows for the incorporation of non-linear effects and random effects using smooth functions and penalized splines. However, its primary emphasis on generalized additive models restricts its suitability for other specialized models, limiting the range of link function combinations available.

In the regression framework, model regularization and variable selection are also essential tasks. These techniques aim to reduce the space of explanatory variables while improving the model estimation and the prediction accuracy. We propose within the functionalities of the **GLMcat** package the conventional stepwise approach. In high-dimensional problems, it is also important to consider regularization methods. For categorical variables, the elastic net penalty can be applied to categorical variables with the **ordinalNet** package, however, it is only designed for three ratios. As future work, we will attempt to define regularization and variable selection methods that are valid for any $(r,\ F,\ Z)$ triplet. We expect, with this extension, more detailed and accurate results, for instance, by means of the Student CDF which is less sensitive to noise variables and thus will improve the variable selection task.

# References


Abreu MNS, Siqueira AL, Cardoso CS, Caiaffa WT (2008). "Ordinal Logistic Regression Models: Application in Quality of Life Studies." *Cadernos De Saude Publica*, **24**(Suppl 4), s581–91. doi:10.1590/s0102-311x2008001600010.

Agresti A (1989). "Tutorial on Modeling Ordered Categorical Response Data." *Psychological Bulletin*, **105**, 290–301. doi:10.1037/0033-2909.105.2.290.

Agresti A (2010). *Analysis of Ordinal Categorical Data*, volume 656. John Wiley & Sons.

Albert A, Anderson JA (1984). "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika*, **71**(1), 1–10. doi:10.1093/biomet/71.1.1.

Ananth CV, Kleinbaum DG (1997). "Regression Models for Ordinal Responses: A Review of Methods and Applications." *International Journal of Epidemiology*, **26**(6), 1323–1333. doi:10.1093/ije/26.6.1323.

Andersen EB (1995). "Polytomous Rasch Models and Their Estimation." In *Rasch Models*, pp. 271–291. Springer-Verlag. doi:10.1007/978-1-4612-4230-7_15.

Armstrong BG, Sloan M (1989). "Ordinal Regression Models for Epidemiologic Data." *American Journal of Epidemiology*, **129**(1), 191–204. doi:10.1093/oxfordjournals.aje.a115109.

Bouscasse H, Joly I, Peyhardi J (2019). "A New Family of Qualitative Choice Models: An Application of Reference Models to Travel Mode Choice." *Transportation Research Part B: Methodological*, **121**, 74–91. doi:10.1016/j.trb.2018.12.010.

Brent RP (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall.

Bürkner PC (2017). "**brms**: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software*, **80**(1), 1–28. doi:10.18637/jss.v080.i01.

Bürkner PC, Vuorre M (2019). "Ordinal Regression Models in Psychology: A Tutorial." *Advances in Methods and Practices in Psychological Science*, **2**(1), 77–101. doi:10.1177/2515245918823199.

Christensen RHB (2024). ***ordinal**: Regression Models for Ordinal Data*. doi:10.32614/CRAN.package.ordinal. R package version 2023.12-4.1.

Coull BA, Agresti A (2000). "Random Effects Modeling of Multiple Binomial Responses Using the Multivariate Binomial Logit-Normal Distribution." *Biometrics*, **56**(1), 73–80. doi:10.1111/j.0006-341x.2000.00073.x.

Cox DR (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society B*, **34**, 187–220. doi:10.1111/j.2517-6161.1972.tb00899.x.

Eddelbuettel D, François R (2011). "**Rcpp**: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**(8), 1–18. doi:10.18637/jss.v040.i08.

Fahrmeir L, Tutz G (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag. doi:10.1007/978-1-4757-3454-6.

Fienberg SE (1980). *The Analysis of Cross-Classified Categorical Data*. MIT press, Massachusetts.

Forinash CV, Koppelman FS (1993). "Application and Interpretation of Nested Logit Models of Intercity Mode Choice." *Transportation Research Record*, **1413**, 98–106.

Geweke J, Keane M, Runkle D (1994). "Alternative Computational Approaches to Inference in the Multinomial Probit Model." *The Review of Economics and Statistics*, **76**(4), 609–632. doi:10.2307/2109766.

Goodman LA (1983). "The Analysis of Dependence in Cross-Classifications Having Ordered Categories, Using Log-Linear Models for Frequencies and Log-Linear Models for Odds." *Biometrics*, **39 1**, 149–60. doi:10.2307/2530815.

Greene WH (2003). *Econometric Analysis*. Pearson Education.

Gutiérrez PA, Pérez-Ortiz M, Sánchez-Monedero J, Fernández-Navarro F, Hervás-Martínez C (2016). "Ordinal Regression Methods: Survey and Experimental Study." *IEEE Transactions on Knowledge and Data Engineering*, **28**(1), 127–146. doi:10.1109/tkde.2015.2457911.

Harrell F (2025). ***rms**: Regression Modeling Strategies*. doi:10.32614/CRAN.package.rms. R package version 8.0-0.

Harrell FE (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer-Verlag.

Hartzel J, Agresti A, Caffo B (2001). "Multinomial Logit Random Effects Models." *Statistical Modelling*, **1**(2), 81–102. doi:10.1177/1471082x0100100201.

Hothorn T (2020). "Most Likely Transformations: The **mlt** Package." *Journal of Statistical Software*, **92**(1), 1–68. doi:10.18637/jss.v092.i01.

Hothorn T, Siegfried S, Kook L (2025). ***tram**: Transformation Models*. doi:10.32614/CRAN.package.tram. R package version 1.2-3.

IBM Corporation (2017). *IBM SPSS Statistics 25*. Armonk. URL http://www.ibm.com/analytics/us/en/technology/spss/.

Johnson NL, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions*. John Wiley & Sons.

Lall R, Campbell M, Walters S, Morgan K, MRC CFAS Team (2002). "A Review of Ordinal Regression Models Applied on Health-Related Quality of Life Assessments." *Statistical Methods in Medical Research*, **11**. doi:10.1191/0962280202sm271ra.

Lange KL, Little RJA, Taylor JMG (1989). "Robust Statistical Modeling Using the $t$ Distribution." *Journal of the American Statistical Association*, **84**(408), 881–896. doi:10.1080/01621459.1989.10478852.

León L, Peyhardi J, Trottier C (2025). ***GLMcat**: Generalized Linear Models for Categorical Responses*. doi:10.32614/CRAN.package.GLMcat. R package version 1.0.0.

Liddell T, Kruschke J (2018). "Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong?" *Journal of Experimental Social Psychology*, **79**, 328–348. doi:10.1016/j.jesp.2018.08.009.

Liu I, Agresti A (2005). "The Analysis of Ordered Categorical Data: An Overview and a Survey of Recent Developments." *Test*, **14**(1), 1–73. doi:10.1007/bf02595397.

Liu X (2009). "Ordinal Regression Analysis: Fitting the Proportional Odds Model Using Stata, SAS and SPSS." *Journal of Modern Applied Statistical Methods*, **8**(2), 30. doi:10.22237/jmasm/1257035340.

Louviere JJ, Hensher DA, Swait JD, Adamowicz W (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge. doi:10.1017/cbo9780511753831.

Lovelace R, Morgan M, Hama L, Padgham M, Ranzolin D, Sparks A (2019). "**stats19**: A Package for Working with Open Road Crash Data." *The Journal of Open Source Software*, **4**(33), 1181. doi:10.21105/joss.01181.

Luce RD (1959). *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons, New York.

Maxwell AE (1961). *Analysing Qualitative Data*. Methuen.

McCullagh P (1980). "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society B*, **42**(2), 109–142. doi:10.1111/j.2517-6161.1980.tb01109.x.

Mellenbergh GJ (1995). "Conceptual Notes on Models for Discrete Polytomous Item Responses." *Applied Psychological Measurement*, **19**(1), 91–100. doi:10.1177/014662169501900110.

O'Connell AA (2006). *Logistic Regression Models for Ordinal Response Variables*, volume 146. Sage. doi:10.4135/9781412984812.

Peyhardi DJ (2020). "Robustness of Student Link Function in Multinomial Choice Models." *Journal of Choice Modelling*, **36**, 100228. doi:10.1016/j.jocm.2020.100228.

Peyhardi J, Trottier C, Guédon Y (2015). "A New Specification of Generalized Linear Models for Categorical Responses." *Biometrika*, **102**(4), 889–906. doi:10.1093/biomet/asv042.

Peyhardi J, Trottier C, Guédon Y (2016). "Partitioned Conditional Generalized Linear Models for Categorical Responses." *Statistical Modelling*, **16**(4), 297–321. doi:10.1177/1471082x16644874.

Rasch G (1961). "On General Laws and the Meaning of Measurement in Psychology." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pp. 321–333.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reinhard R, Rutrecht HM, Hengstenberg P, Tutulmaz E, Geissler B, Hecht H, Muttray A (2017). "The Best Way to Assess Visually Induced Motion Sickness in a Fixed-Base Driving Simulator." *Transportation Research Part F: Traffic Psychology and Behaviour*, **48**, 74–88. doi:10.1016/j.trf.2017.05.005.

Ripley BD, Venables WN (2025). ***nnet**: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. doi:10.32614/CRAN.package.nnet. R package version 7.3-20.

Ripley BD, Venables WN, Bates DM, Hornik K, Gebhardt A, Firth D (2025). ***MASS**: Support Functions and Datasets for Venables and Ripley's MASS*. doi:10.32614/CRAN.package.MASS. R package version 7.3-65.

Sankey SS, Weissfeld LA (1998). "A Study of the Effect of Dichotomizing Ordinal Data Upon Modeling." *Communications in Statistics-Simulation and Computation*, **27**(4), 871–887. doi:10.1080/03610919808813515.

SAS Institute Inc (2020). *SAS/STAT Software, Version 9.4*. Cary. URL http://www.sas.com/.

Scott SC, Goldberg MS, Mayo NE (1997). "Statistical Assessment of Ordinal Outcomes in Comparative Studies." *Journal of Clinical Epidemiology*, **50**(1), 45–55. doi:10.1016/s0895-4356(96)00312-5.

StataCorp (2015). *Stata Statistical Software: Release 14*. College Station. URL http://www.stata.com/.

Stroustrup B (2013). *The C++ Programming Language*. 4th edition. Addison-Wesley.

Tutz G (1991). "Sequential Models in Categorical Regression." *Computational Statistics & Data Analysis*, **11**(3), 275–295. doi:10.1016/0167-9473(91)90086-h.

Tutz G (2011). *Regression for Categorical Data*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. doi:10.1017/cbo9780511842061.

Tutz G, Hennevogl W (1996). "Random Effects in Ordinal Regression Models." *Computational Statistics & Data Analysis*, **22**(5), 537–557. doi:10.1016/0167-9473(96)00004-7.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York. doi:10.1007/978-0-387-21706-2.

Wickham H (2015). *R Packages: Organize, Test, Document, and Share Your Code*. O'Reilly Media.

Wood SN (2011). "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society B*, **73**(1), 3–36. doi:10.1111/j.1467-9868.2010.00749.x.

Wooldridge JM (2012). *Introductory Econometrics: A Modern Approach*. 5th edition. Cengage Learning.

Wurm M, Rathouz P, Hanlon B (2025). ***ordinalNet**: Penalized Ordinal Regression*. doi:10.32614/CRAN.package.ordinalNet. R package version 2.13.

Zhang Q, Ip EH (2012). "Generalized Linear Model for Partially Ordered Data." *Statistics in Medicine*, **31**(1), 56–68. doi:10.1002/sim.4318.

# A. CDF of the non-central $t$ distribution

$$F_{\nu,\mu}(\eta) = \Phi(-\mu) + \frac{1}{2}\sum_{j=0}^{\infty}\left(p_j I_y\left(j+\frac{1}{2},\frac{\nu}{2}\right) + q_j I_y\left(j+1,\frac{1}{2}\right)\right),$$

where:

- $\Phi$ is the CDF of the standard normal distribution,
- $I_y(a,b)$ is the regularized incomplete beta function,
- $y = \dfrac{\eta^2}{\eta^2+\nu}$,
- $p_j = \exp\left(-\dfrac{\mu^2}{2}\right)\left(\dfrac{\mu^2}{2}\right)^j$, and
- $q_j = \dfrac{\mu}{\sqrt{2}\Gamma(j+3/2)}\exp\left(-\dfrac{\mu^2}{2}\right)\left(\dfrac{\mu^2}{2}\right)^j$.

# B. Jacobian matrices

The Jacobian matrices $\partial\boldsymbol{\pi}/\partial\boldsymbol{r}$ used as part of the Fisher's scoring algorithm are presented for each ratio.

*Cumulative:*

$$\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{r}} = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ & & & & 1 \end{pmatrix}.$$

In the following, we present the form of the element corresponding to row $i$ and column $j$ of the Jacobian Matrix.

*Adjacent:*

$$\frac{\partial\pi_j}{\partial r_i} = \frac{1}{F(\eta_i)[1-F(\eta_i)]}\begin{cases}\pi_j(1-\gamma_i) & \text{if } i \geq j,\\ -\pi_j\gamma_i & \text{otherwise,}\end{cases}$$

where $\gamma_i = \mathsf{P}(Y \leq i) = \sum_{k=1}^{i}\pi_k$.

*Sequential:*

$$\frac{\partial\pi_j}{\partial r_i} = \begin{cases}\prod_{k=1}^{j-1}\{1-F(\eta_k)\} & \text{if } i = j,\\ -F(\eta_j)\prod_{k=1,k\neq i}^{j-1}\{1-F(\eta_k)\} & \text{if } i < j,\\ 0 & \text{otherwise.}\end{cases}$$

*Reference:*

$$\frac{\partial \pi_j}{\partial r_i} = \frac{\mathsf{COV}(Y_i, Y_j)}{F(\eta_i)[1 - F(\eta_i)]}.$$

Refer to the Supplementary Material of Peyhardi *et al.* (2015) for further details.

# C. Proofs

## C.1. Proof of Proposition 1

Consider the distribution of $Y$ defined by the $(adjacent, F, Z)$ model. The adjacent ratio for category $J - j$ can be written as

$$r_{(J-j)}(\boldsymbol{\pi}) = \frac{\pi_{J-j}}{\pi_{J-j} + \pi_{J-j+1}} \tag{6}$$

for all $j \in \{1, \ldots, J-1\}$. Simultaneously, consider the distribution of $\tilde{Y}$ defined by the $(adjacent, F, Z)_{\tilde{\sigma}}$ model (equivalent to $r(\boldsymbol{\pi}_\sigma) = F(Z\beta)$), where $\tilde{\sigma}$ is the reverse permutation, i.e., $\tilde{\sigma}(j) = J - j + 1$ for all $j \in \{1, \ldots, J-1\}$, we can prove the next equality

$$r_j(\boldsymbol{\pi}_{\tilde{\sigma}}) = 1 - r_{\tilde{\sigma}(j+1)}(\boldsymbol{\pi}) \quad (j = 1, \ldots, J-1) \tag{7}$$

through the ratio expressed for element $\tilde{\sigma}(j+1) = J - j$ in Equation 6 where

$$\begin{aligned} 1 - r_{\tilde{\sigma}(j+1)}(\boldsymbol{\pi}) &= \frac{\pi_{J-j+1}}{\pi_{J-j} + \pi_{J-j+1}} \\ &= \frac{\pi_{\sigma(j)}}{\pi_{\sigma(j+1)} + \pi_{\sigma(j)}} \\ &= r_j(\boldsymbol{\pi}_{\tilde{\sigma}}). \end{aligned}$$

Given that $r_j(\boldsymbol{\pi}_{\tilde{\sigma}}) = F(\eta_j)$ and using Equation 7, we obtain that $r_{J-j}(\boldsymbol{\pi}) = \tilde{F}(-\eta_j)$. If we denote $i = J - j$ the last equality becomes

$$r_i(\boldsymbol{\pi}) = \tilde{F}(-\eta_{J-i})$$

for all $j \in \{1, \ldots, J-1\}$. Hence $\tilde{Y}$ follows the $(adjacent, \tilde{F}, -PZ)$ model, where $P$ is the restricted reverse permutation matrix of dimension $J - 1$ defined in Equation 4. Since $P$ has full rank, the design matrices $Z$ and $-PZ$ are equivalent, meaning the $(adjacent, F, Z)$ model is equal to the $(adjacent, \tilde{F}, -PZ)$ model. The above can be similarly demonstrated for the cumulative ratio, but not for the sequential ratio given that Equation 7 is invalid for these models. To prove it by contradiction, the reader can assume that the statement is false, proceed from there, and at some point, a contradiction will result.

## C.2. Proof of equality of $(reference, logistic, AZ)$ and $(adjacent, logistic, Z)$

Assume that the distribution of $Y$ is defined by the $(reference, logistic, Z)$ model. For $j = 1, \ldots, J$ we obtain

$$\ln\left(\frac{\pi_j}{\pi_J}\right) = \eta_j.$$

The adjacent ratio can be rewritten in terms of the reference ratio since

$$\ln\left(\frac{\pi_j}{\pi_{j+1}}\right) = \ln\left(\frac{\pi_j}{\pi_J}\right) + \ln\left(\frac{\pi_j}{\pi_{j+1}}\right),$$

therefore, using the reparametrization

$$= \begin{cases} \eta'_j = \eta_j - \eta_{j+1}, \text{ for } j = 1, \ldots, J-2, \\ \eta'_{J-1} = \eta_{J-1} \end{cases}$$

represented by the matrix

$$A = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ & & & & 1 \end{pmatrix}$$

of dimension $J-1$, we obtain the equality $(\textit{reference}, \textit{logistic}, AZ) = (\textit{adjacent}, \textit{logistic}, Z)$, and, given that $A$ is invertible, we obtain $(\textit{reference}, \textit{logistic}, Z) = (\textit{adjacent}, \textit{logistic}, A^{-1}Z)$.

**Affiliation:**

Lorena León
CIRAD, UMR AGAP Institut
IMAG, Université de Montpellier
34000 Montpellier, France
E-mail: ylorenaleonv@gmail.com

Jean Peyhardi
IMAG, CNRS, Université de Montpellier
34000 Montpellier, France
E-mail: jean.peyhardi@umontpellier.fr

Catherine Trottier
IMAG, CNRS, Université de Montpellier
AMIS, Université Paul Valéry Montpellier 3
34000 Montpellier, France
E-mail: catherine.trottier@umontpellier.fr

*Journal of Statistical Software* https://www.jstatsoft.org/
published by the Foundation for Open Access Statistics https://www.foastat.org/
September 2025, Volume 114, Issue 9 *Submitted:* 2024-01-19
doi:10.18637/jss.v114.i09 *Accepted:* 2024-09-24