Ordinal regression models made easy. A tutorial on parameter interpretation, data simulation, and power analysis.

Filippo Gambarota[1] & Gianmarco Altoè[1]

[1] Department of Developmental Psychology and Socialization, University of Padova, Italy

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. Filippo Gambarota: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Gianmarco Altoè: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Filippo Gambarota, Postal address. E-mail: filippo.gambarota@unipd.it

Abstract

Abstract here

Keywords: ordinal, likert, simulations, power

Word count: X

Ordinal regression models made easy. A tutorial on parameter interpretation, data simulation, and power analysis.

Introduction

Risolvere un pochino la nomenclatura

https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf

> The name, cumulative link models is adopted from Agresti (2002), but the model class has been referred to by several other names in the literature, such as ordered logit models and ordered probit models (Greene and Hensher 2010) for the logit and probit link functions. The cumulative link model with a logit link is widely known as the proportional odds model due to McCullagh (1980) and with a complementary log-log link, the model is sometimes referred to as the proportional hazards model for grouped survival times.

Da citare in qualche modo Kemp and Grace (2021)

> "As a matter of fact, most of the scales used widely and effectively by psychologists are ordinal scales. In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. On the other hand for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results." (Stevens, 1946, p. 679.)

> It is straightforward to construct ordinal scales that do not involve rank ordering. For example, one can take the first element encountered and arbitrarily assign it the number 100. If the next element encountered is smaller it is given a smaller arbitrary number, 53 say. If the third element is between

> these two, it can be given the number 86, and so on. If this construction method is used, the difference between the elements assigned the numbers 70 and 80 will not in any important sense be equal to the difference between the elements assigned 90 and 100, and the intervals between the numbers are not really interpretable. Note, too, that monotonic transformations of the scale essentially leave the mea- sure unaffected.

interessante come Cliff (2016) descriva che la maggiorparte dellle domande di ricerca in psicologia siano in riferimento alla location di un costrutto e possano essere gestite in modo ordinale

The use of ordinal data is widespread in Psychology. Usually items from questionnaires are created using Likert scales where a certain psychological traits is measured using an intuitive scale with 3-5 or more anchor points. These measure cannot be considered *metric* measure on interval or ratio scales [stevens] but the categories are ordered.

Other examples here

Despite the usage of ordinal variables, statistical models made for these type of data are rarely used in Psychology. Liddell and Kruschke (2018) reported that the majority of published papers using likert-like measures used standard methods to analyze the data. In practical terms, they use *metric* models where the response variable cannot be considered fully numeric (vedi se c'è un termine tipo scala a rapporti).

altro sui modelli

Theoretically, using a *metric* for ordinal data is not appropriate but understanding the actual impact is not straightforward. Liddell and Kruschke (2018) did a comprehensive work about pitfalls of analzying ordinal data as metric. They showed that metric models produce higher type-1 and type-2 errors compared to the ordinal models. In particular, given the bounded nature of ordinal data, difference in the underlying latent distribution are not always captured by the metric model that simply estimate the mean of the ordinal variable. This is even more relevant when the underluing variance of the ordinal variables are not homogeneous. Furthermore they presented some situations where the metric model could be wrong in the opposite direction, finding an effect with the wrong sign (type-s error).

Finally, they also demonstrated that even in the best condition where e.g. comparing two groups the variance are equal, the ordinal model is more powerful than the metric model given the underestimation of the true effect size from the latter.

> Some authors have argued that, despite the ordinal character of individual Likert items, averaged ordinal items can have an emergent property of an interval scale and so it is appropriate to apply metric methods to the averaged values (e.g., Carifio & Perla, 2007, 2008).

vedi come gestire questo

anche questo è importante > Ordered-probit models typically assume that the thresholds (k) are the same across all groups because the thresholds are theoretically linked to the response measure, not to the predictor value. For example, when asked, “How happy are you?” with response options ‘1’ = very unhappy, ‘2’ = mildly unhappy, ‘3’ = neutral, ‘4’ = mildly happy, ‘5’ = very happy,” the latent thresholds between ordinal levels are assumed to be implicit in the phrasing of the question, regardless of other aspects of the respondent or situation. In other words, the thresholds are assumed to be part of the measurement procedure, not dependent on the value of the predictor or covariate. This can be technically referred to as a type of measurement invariance.

Da citare come paper iniziale sui modelli (McCullagh, 1980).

questo per la tassonomia dei diversi modelli ordinali (Tutz, 2022)

da vedere magari per una critica tipo Gomila sul binary model (Robitzsch, 2020)

Metriv vs ordinal models

- the main difference is that the metric model assign a number to each label of the discrete variable assuming that the distance is the same. (Liddell & Kruschke, 2018)

lm on latent vs ordinal

as suggested https://people.vcu.edu/~dbandyop/BIOS625/CLM_R.pdf, running a lm on the latent variables gives similar parameter as the clm. Of course, we are able to do this with real data given that the ordinal variable is the observed version of an unobserved latent variable. But in simulation this is useful to understand what the cumulative model is doing.

```
##
## Call:
## lm(formula = ys ~ x1 + x2, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0126 -0.6777 -0.0003  0.6754  4.1015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.003164   0.003161  -1.001    0.317
## x1           1.005849   0.003146 319.749   <2e-16 ***
## x2           0.498501   0.003156 157.971   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9996 on 99997 degrees of freedom
## Multiple R-squared:  0.5589, Adjusted R-squared:  0.5589
## F-statistic: 6.335e+04 on 2 and 99997 DF,  p-value: < 2.2e-16
```

```
## formula: y ~ x1 + x2
## data:    dat
##
##  link   threshold nobs  logLik     AIC       niter max.grad cond.H
##  probit flexible  1e+05 -122018.81 244049.62 5(0)  9.59e-08 9.1e+00
##
## Coefficients:
##     x1     x2
## 1.0064 0.4993
##
## Threshold coefficients:
##     1|2     2|3     3|4     4|5
## -0.8347 -0.2504  0.2529  0.8444
```

```
##  (Intercept)           x1           x2
## -0.003165429  1.006273800  0.498711694
```

Kruscke parametrization

Liddell and Kruschke (2018) and Kruschke (2015) proposed an alternative parametrization to understand the model parameters. They used a probit model where thresholds and regression parameters are estimated on the scale of the ordinal variable compared to standard ordinal regression where they refers to the quantile of the latent variables for the threshold and z score or odds ratio for the regression coefficients. They implemented the model in Jags and provided some equations and R functions to convert from the standard parametrization to the proposed one.

Odds ratios and cumulative odds ratios

- https://online.stat.psu.edu/stat504/lesson/4/4.1

In this section we introduce the main concept to understand the effects of a logit model.

Demonstrating the proportional odds assumption

basically given that the effect of β is constant the odds ratio is independent from the thresholds α_j (Liu, He, Tu, & Tang, 2023). In the current paper we are simulating data using a single set of βs thus the proportional odds assumption is true. Liu et al. (2023) reviewed the available methods to test this assumption.

- https://hbiostat.org/ordinal/impactpo.pdf blog su proportional odds, mi pare di capire che non sia così problematica come assunzione

sarebbe da capire quali sono i rischi di assumere questo ed eventualmente se simulare non proportional odds è troppo complicato.

Let's simulate the effect of a binary predictor on ordinal scale 1-5:

```r
b1 <- log(3) # log odds ratio
n <- 1e4
x <- rep(c("a", "b"), each = n/2)
dat <- data.frame(x = x)
probs <- rep(1/5, 5) # for the group "a", uniform probabilities
dat <- sim_ord_latent(~x, By = b1, probs = probs, data = dat, link = "logit")
```

```
fit <- clm(y ~ x, data = dat, link = "logit")
pr <- predict(fit, data.frame(x = unique(x)))$fit
pr
```

```
##            1         2         3         4         5
## 1 0.19720971 0.2039506 0.2009355 0.1973976 0.2005066
## 2 0.07341871 0.1042636 0.1503074 0.2345890 0.4374213
```

Basically the proportional odds suggest that:

$$log(\frac{P(y \leq 1)}{P(y > 1)})$$

Is the same regardless the level of the x predictor. Thus:

$$\log\left(\frac{\frac{P(y \leq 1|x_0)}{P(y > 1|x_0)}}{\frac{P(y \leq 2|x_0)}{P(y > 2|x_0)}}\right) = \log\left(\frac{\frac{P(y \leq 1|x_1)}{P(y > 1|x_1)}}{\frac{P(y \leq 2|x_1)}{P(y > 2|x_1)}}\right)$$

```
a_or1vs2345 <- filor::odds(pr[1, 1]) # 1 vs 2 3 4 5
a_or12vs345 <- filor::odds(sum(pr[1, 1:2])) # 1 vs 2 3 4 5


b_or1vs2345 <- filor::odds(pr[2, 1]) # 1 vs 2 3 4 5
b_or12vs345 <- filor::odds(sum(pr[2, 1:2])) # 1 vs 2 3 4 5


c(xa = log(a_or1vs2345 / a_or12vs345), xb = log(b_or1vs2345 / b_or12vs345))
```

```
##        xa        xb
## -1.003193 -1.003193
```

Proportional odds can be also visualized by plotting the cumulative probabilities of y, in terms of $g(P(y \leq j))$ (where $g()$ is the logit link function) as a function of the predictor

x. If the proportional odds assumptions holds the slopes are parallel (also known as parallel regression assumption). The Figure 1 depicts the assumption of proportional odds in the probability and logit space.

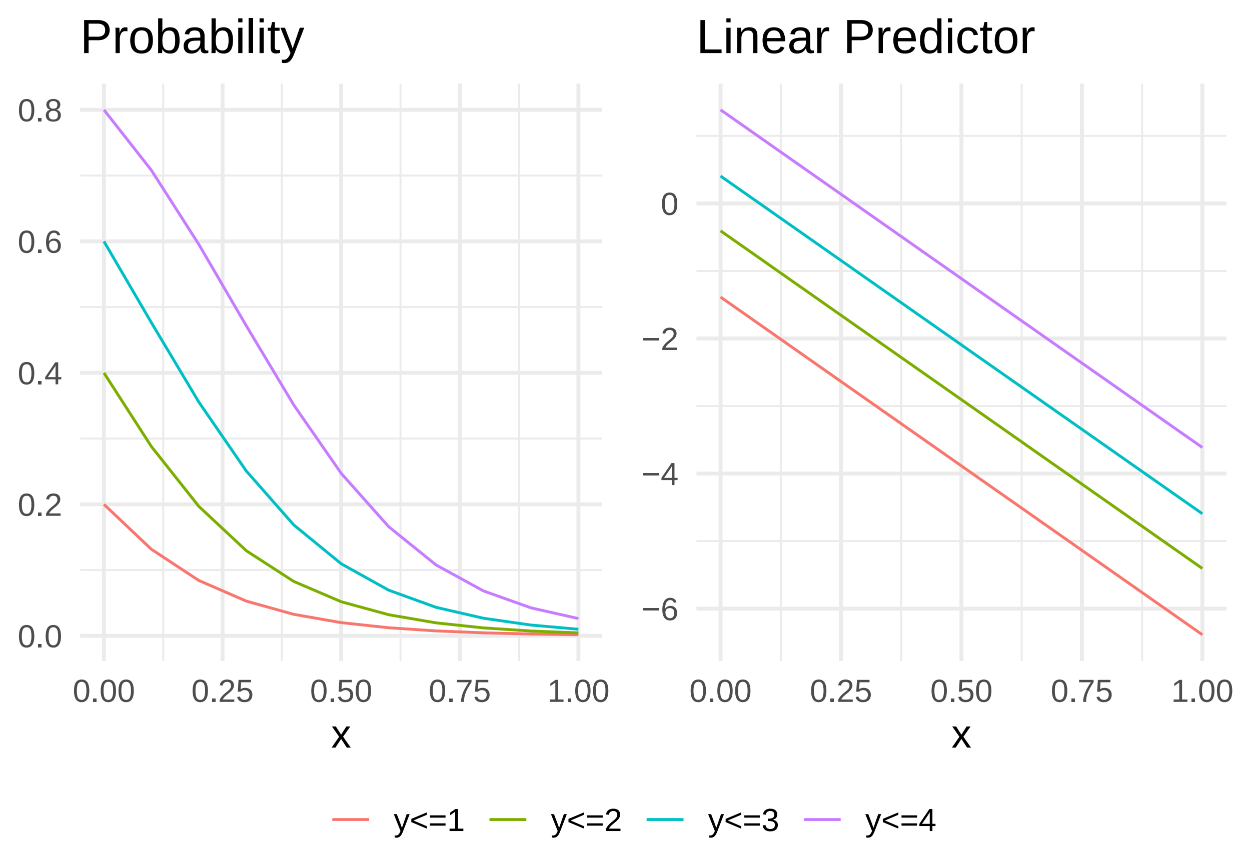

Figure 1. POA assumption

Harrell (2015) proposed a very intuitive plot to assess the proportional odds assumption and eventually the degree of deviation from the ideal case. Basically predictor is plotted against the logit of the cumulative probability. Distances between pairs of symbols should be similar across levels of the predictors. Numerical predictors can be binned before plotting the corresponding logit. Figure 2 depicts an example with simulated data satisfying the proportional odds assumption.

From a statistical point of view, the proportional odds assumption can be assessed by fitting $k-1$ binomial regressions and checking if the estimated β is similar between

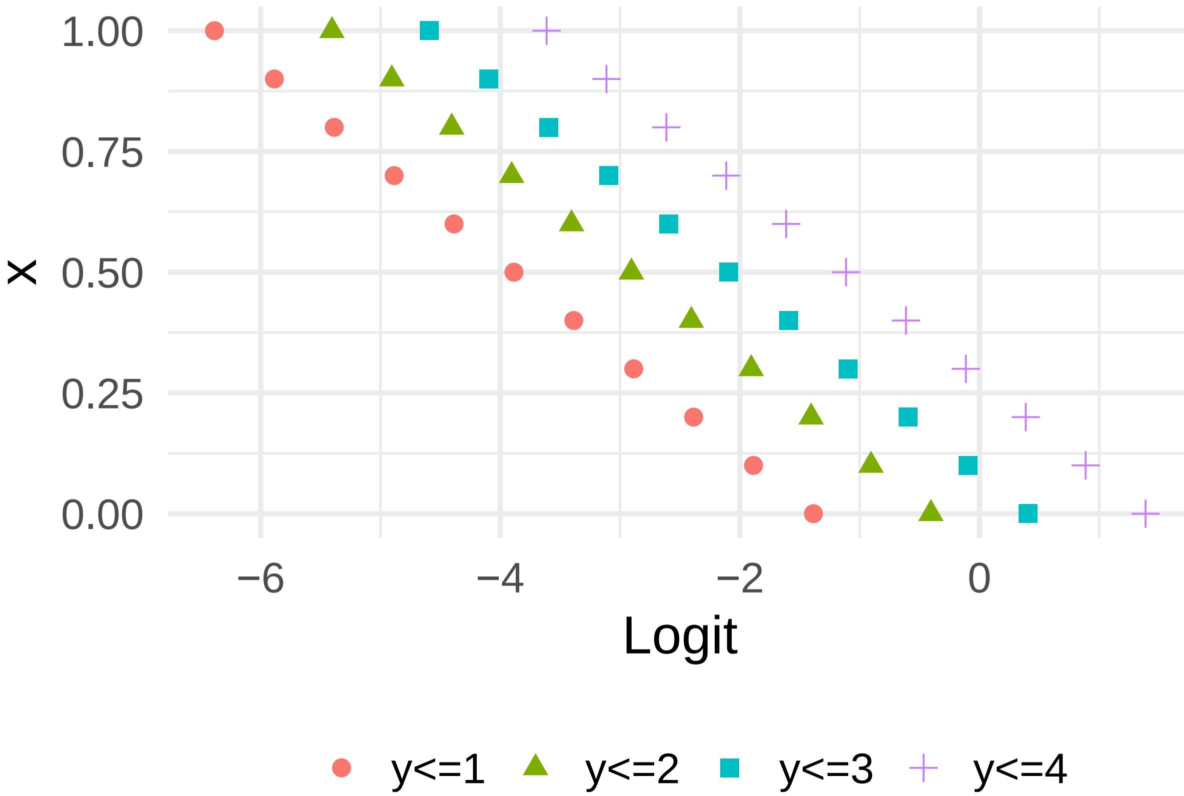

Figure 2. Harrel Poa Plot

regressions. The regressions are estimated by creating $k-1$ dummy variables from the ordinal y. The code below show that simulating data with the proportional odds create $k-1$ binomial regression with similar βs.

```
k <- 4
n <- 1e5
dat <- data.frame(
  x = runif(n)
)

dat <- sim_ord_latent(~x, By = 5, probs = rep(1/k, k), data = dat, link = "logit")

# create dummy variables
```

```r
dat$y1vs234 <- ifelse(dat$y <= 1, 1, 0)
dat$y12vs34 <- ifelse(dat$y <= 2, 1, 0)
dat$y123vs4 <- ifelse(dat$y <= 3, 1, 0)

fit1vs234 <- glm(y1vs234 ~ x, data = dat, family = binomial(link = "logit"))
fit12vs34 <- glm(y12vs34 ~ x, data = dat, family = binomial(link = "logit"))
fit123vs4 <- glm(y123vs4 ~ x, data = dat, family = binomial(link = "logit"))

car::compareCoefs(fit123vs4, fit12vs34, fit1vs234)
```

```
## Calls:
## 1: glm(formula = y123vs4 ~ x, family = binomial(link = "logit"), data = dat)
## 2: glm(formula = y12vs34 ~ x, family = binomial(link = "logit"), data = dat)
## 3: glm(formula = y1vs234 ~ x, family = binomial(link = "logit"), data = dat)
##
##             Model 1 Model 2 Model 3
## (Intercept)  1.1117  0.0148 -1.1014
## SE           0.0151  0.0163  0.0216
##
## x           -5.0263 -5.0329 -4.9763
## SE           0.0365  0.0486  0.0732
##
```

Logit vs Probit model

When fitting an ordinal regression the two mostly used link functions are the *probit* and *logit*. From the distribution point of view the two functions are very similar. The *probit* model is based on a cumulative Gaussian distribution while the *logit* model is based on a logistic distribution. Figure 3 depict the two cumulative distributions.

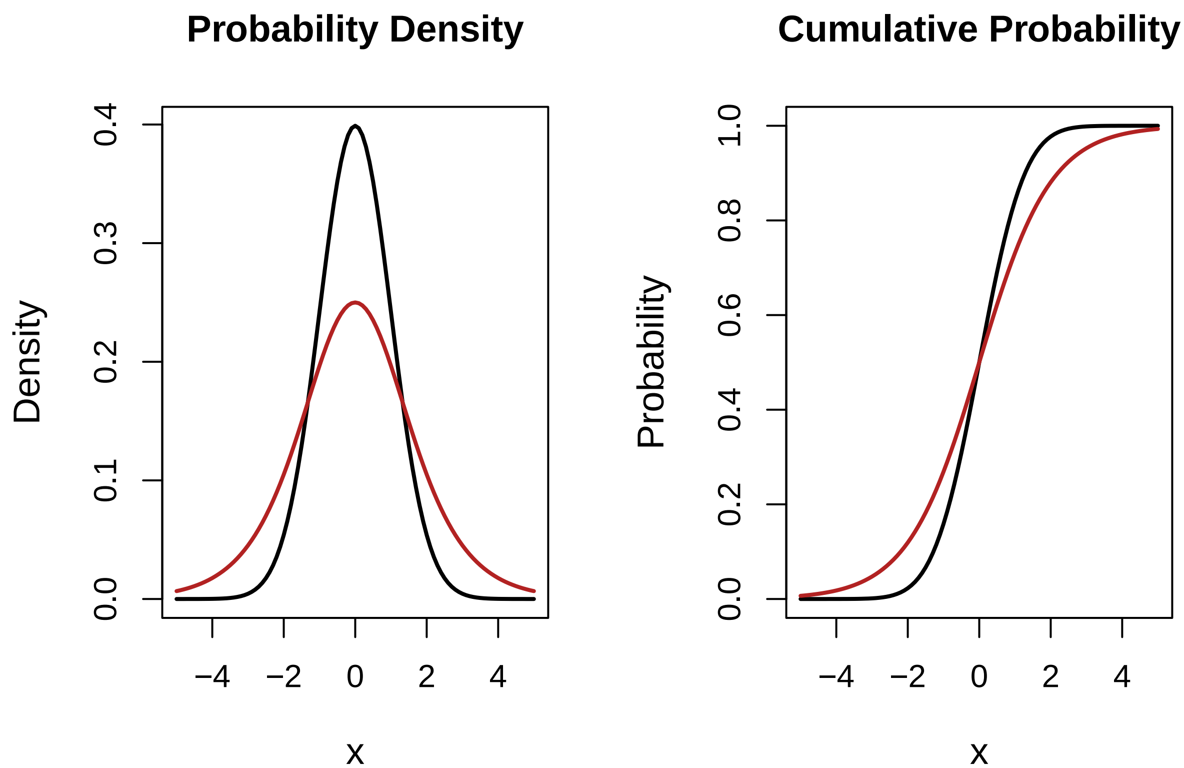

Figure 3. Logit vs Probit

Given the different underlying distribution, the parameters have a different interpretation under the two models. The probit model assume a latent standard normal distribution with and $\sigma = 1$. The logit model assume a logistic distribution with $\sigma^2 = \pi^2/3$. Thus regression coefficients β represents the increase in standard deviations units. The interpretation in terms of latent distribution is particularly useful for the probit model where βs can be interpreted in a Cohen's d like manner. Furthermore, there is the possibility to directly map parameters from signal detection theory (Green, Swets, & Others, 1966; Stanislaw & Todorov, 1999) into an ordinal probit regression. In practical terms, the thresholds are the criterion cut-offs and the β_1 is the d′ parameter (DeCarlo, 1998; Knoblauch & Maloney, 2012).

The latent formulation of the ordinal regression model allow to think in standard linear regression terms for choosing and interpreting model parameters before converting into the probability space. Both the standard Normal and the logistic distribution are defined with a location (μ) and a scale (s) parameter. For the Normal distribution, μ is the mean and scale is the variance (σ^2). For the logistic distribution the variance is $\sigma^2 = \frac{s^2\pi^2}{3}$. Thus fixing μ and $\sigma^2 = 1$ (the default), the two distributions are similar with the logistic having more variance. For this reason, the latent formulation for parameter intepretation is particularly useful for the probit model because the β is directly interpreted in standard deviaton units that by default are fixed to 1. With a binary predictor x, β is the shift of the latent distribution increasing x by one unit. This can be directly interpreted as a standardized mean difference effect size (e.g., Cohen's d). This is the same for the logistic distribution but a unit increase in x shift the latent distribution by $\sigma = \sqrt{\frac{s^2\pi^2}{3}} = \frac{s\pi}{\sqrt{3}}$.[1].

Simulating data

There are mainly two ways to simulate data. The first method concerns simulating starting from the latent formulation of the ordinal model. Basically we can simulate the underlying latent distribution and then fixing the thresholds converting the latent continuous variable into the observed ordinal variable.

Scale Effects

The default ordinal regression model assume that the variance of th underlying latent distribution is the same across condition. This is similar to a standard linear regression assuming homogeneity of variance. For example, when comparing two groups or conditions we can run a standard linear model (i.e., a t-test) assuming homogeneity of variances or using the Welch t-test (see Delacre, Lakens, & Leys, 2017). In addition, there are the so-called location-scale models that allows to include predictors also for the scale (e.g., the variance) of the distribution. This can be done also in ordinal regression where instead of assuming the same variance between conditions, the linear predictors can be included.

[1] In the case of a standard logistic distribution ($s^2 = 1$), the standard deviation is $\frac{\pi}{\sqrt{3}}$

Figure ?? depict an example of a comparison between two groups where the two underlyning latent distributions have unequal variance.

Visualizing effects

Before starting with the actual simulation it is useful to plot the predicted probabilities of the response variable y as a function of the predictors. The `cat_latent_plot()` and the `num_latent_plot()` functions are able to visualize the effect of a single categorical or numerical predictor. Specifying the mean and standard deviations of the latent variable in each condition (for the categorical version) or the β (for the numerical version) the function compute the expected probability for each level of the ordinal variable y and visualize the result. In this way the β can be choose in a meaningful way.

```
cat_latent_plot <- function(m = 0,
                            s = 1,
                            th = NULL,
                            probs = NULL,
                            link = c("logit", "probit"),
                            plot = c("probs", "latent", "both")){
  plot <- match.arg(plot)
  lf <- get_link(link)
  x <- paste0("g", 1:length(m))
  lat <- data.frame(
    x = x,
    m = m,
    s = s
  )
```

```r
if(is.null(th)){
  th <- probs_to_th(probs, link)
}

thl <- latex2exp::TeX(sprintf("$\\alpha_{%s}$", 1:length(th)))

if(link == "logit"){
  dfun <- distributional::dist_logistic
  s <- sqrt(vlogit(s))
  title <- "Logistic Distribution"
}else{
  dfun <- distributional::dist_normal
  s <- lat$s
  title <- "Normal Distribution"
}
lat$dist <- dfun(lat$m, lat$s)
range_y <- c(min(lat$m) - max(s) * 5, max(lat$m) + max(s) * 5)

lat_plot <- ggplot(lat,
       aes(x = factor(x), y = m, dist = dist)) +
  ggdist::stat_halfeye(aes(fill = after_stat(factor(findInterval(y, th) + 1))),
                       alpha = 0.85) +
  ylim(range_y) +
  geom_hline(yintercept = th, linetype = "dashed", col = "black", alpha = 0.7) +
  geom_line(aes(x = factor(x), y = m, group = 1)) +
  annotate("label", x = 0.7, y = th, label = thl, size = 5) +
  theme_minimal(15) +
```

```
  theme(legend.position = "bottom",
        legend.title = element_blank(),
        axis.title.x = element_blank()) +
  ylab(latex2exp::TeX("\\mu")) +
  ggtitle(title)



th <- c(-Inf, th, Inf)
lat_ms <- lat[, c("m", "s")]
ps <- apply(lat_ms, 1, function(x) data.frame(t(diff(lf$pfun(th, x[1], x[2])))), simp
ps <- do.call(rbind, ps)
names(ps) <- paste0("y", 1:ncol(ps))
latl <- cbind(lat, ps)
latl <- pivot_longer(latl, starts_with("y"), names_to = "y", values_to = "value")
probs_plot <- ggplot(latl, aes(x = factor(x), y = value, fill = y)) +
  geom_col(position = position_dodge()) +
  theme_minimal(15) +
  theme(legend.position = "bottom",
        legend.title = element_blank(),
        axis.title.x = element_blank()) +
  ylab("Probability")
if(plot == "probs"){
  probs_plot
}else if(plot == "latent"){
  lat_plot
}else{
  legend_b <- get_legend(
```

```
      probs_plot
    )
    plts <- plot_grid(
      lat_plot + theme(legend.position = "none"),
      probs_plot + theme(legend.position = "none")
    )
    plot_grid(plts, legend_b, ncol = 1, rel_heights = c(1, .1))
  }
}
num_latent_plot <- function(x,
                            b1,
                            th = NULL,
                            probs = NULL,
                            link = c("logit", "probit"),
                            nsample = 1e3,
                            size = 20){
  lf <- get_link(link)
  if(is.null(th)){
    th <- probs_to_th(probs, link)
  }
  data <- data.frame(x = seq(min(x), max(x), length.out = nsample))
  data <- get_probs(~x, b1, probs, data, link, append = TRUE)
  datal <- tidyr::pivot_longer(data, starts_with("y"), names_to = "y", values_to = "valu
  ggplot(datal, aes(x = x, y = value, color = y)) +
    geom_line() +
    theme_minimal(size) +
    theme(legend.position = "bottom",
```

```
        legend.title = element_blank()) +
    ylab("Probability") +
    ggtitle(latex2exp::TeX(sprintf("$\\beta_1 = %s$", b1)))
}
```

For example, the Figure 4 depicts the effect of a categorical predictor on a 5-level y assuming uniform probabilities for the reference level and a mean difference on a standardized Normal distribution of 1. The Figure 5 depicts the effect of a continuous predictor x sampled from a standard Normal distribution assuming a $\beta = 1$.

Disclaimer about the functions

The current paper proposed a simplified way with some functions to generate ordinal data. For more complex simulations such as simulating correlated ordinal data the `simstudy` package https://kgoldfeld.github.io/simstudy/articles/ordinal.html proposed a very comprehensive set of data simulation function also for ordinal data.

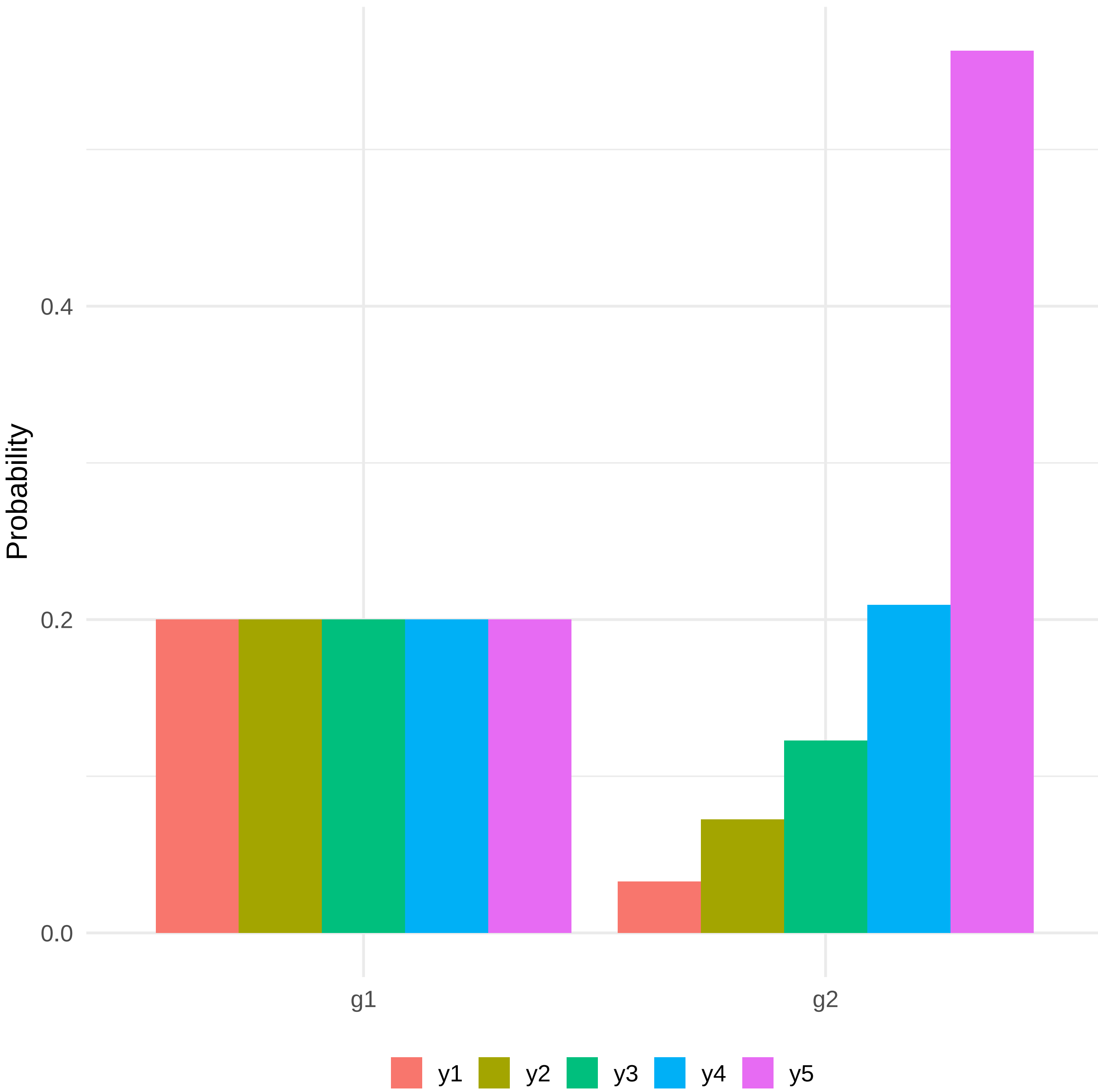

Figure 4. Example categorical

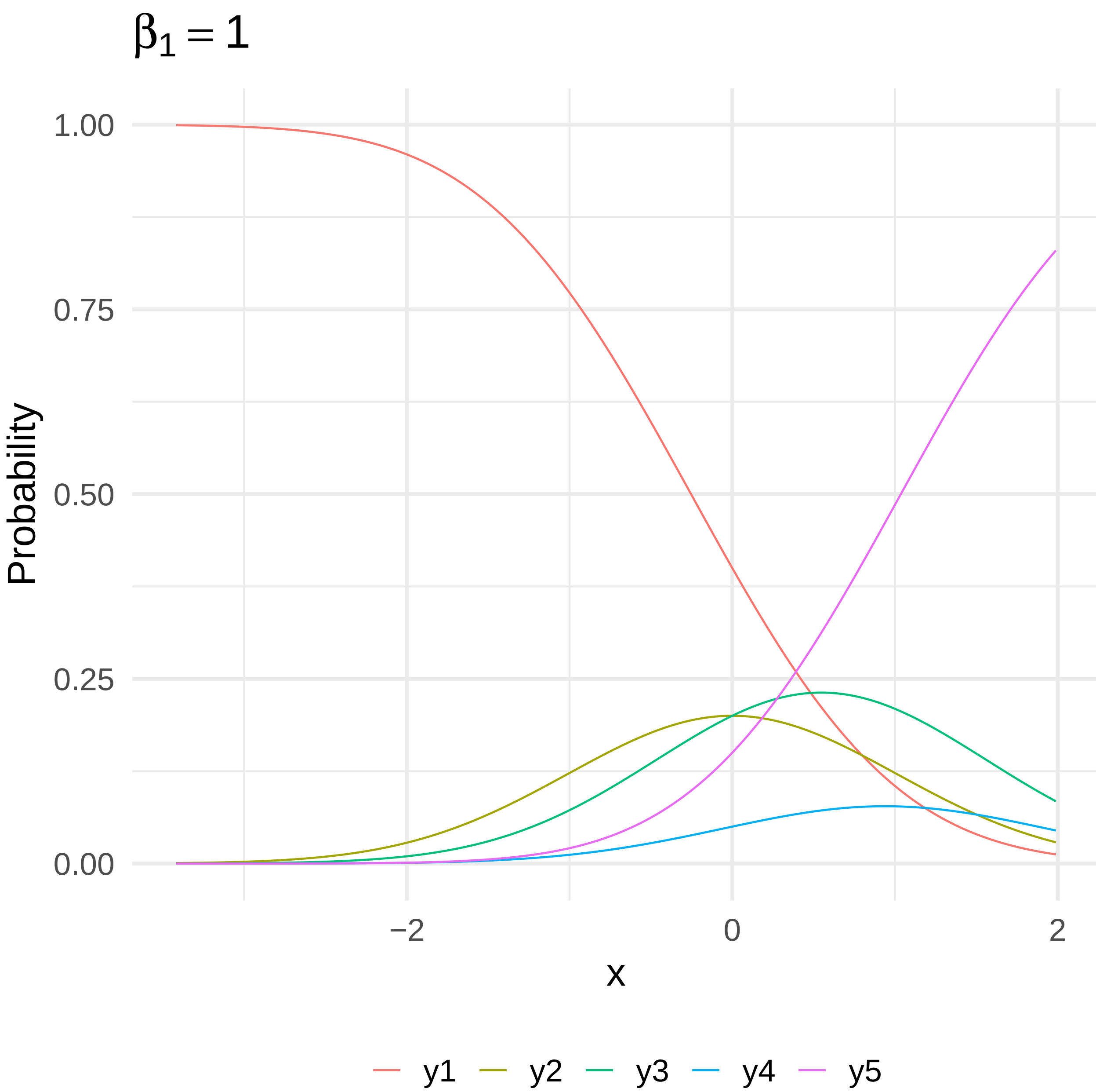

Figure 5. Example numerical

References

Cliff, N. (2016). *Ordinal methods for behavioral data analysis*. London, England: Psychology Press.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186–205. https://doi.org/10.1037/1082-989X.3.2.186

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch's *t*-test instead of student's *t*-test. *International Review of Social Psychology*, *30*, 92. https://doi.org/10.5334/irsp.82

Green, D. M., Swets, J. A., & Others. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. xi, 455–xi, 455). Oxford, England: Wiley New York.

Harrell, F. E. (2015). *Regression modeling strategies*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7

Kemp, S., & Grace, R. C. (2021). Using ordinal scales in psychology. *Methods in Psychology (Online)*, *5*, 100054. https://doi.org/10.1016/j.metip.2021.100054

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in r*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-4475-6

Kruschke, J. K. (2015). *Doing bayesian data analysis*. https://doi.org/10.1016/c2012-0-00477-2

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

Liu, A., He, H., Tu, X. M., & Tang, W. (2023). On testing proportional odds assumptions for proportional odds models. *General Psychiatry*, *36*, e101048. https://doi.org/10.1136/gpsych-2023-101048

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, *42*, 109–127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*, *5*, 589965. https://doi.org/10.3389/feduc.2020.589965

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *31*, 137–149. https://doi.org/10.3758/bf03207704

Tutz, G. (2022). Ordinal regression: A review and a taxonomy of models. *Wiley Interdisciplinary Reviews. Computational Statistics*, *14*. https://doi.org/10.1002/wics.1545