

Resources for Ordinal Regression Models

Frank Harrell

2022-04-27

1 Background

Consider an ordinal outcome variable Y and note that an interval-scaled Y is also ordinal. Such response variables can be analyzed with nonparametric tests and correlation coefficients and with semiparametric ordinal models. Most semiparametric models in use involve modeling cumulative probabilities of Y . Resources here deal primarily with the cumulative probability family of semiparametric models¹.

Semiparametric models are regression models having an intercept for each distinct value of Y , less one. They have the following properties:

- By encoding the entire empirical cumulative distribution of Y in the intercepts, the model assumes nothing about the shape of the Y distribution for any given covariate setting. The distribution can be continuous, discontinuous, have arbitrary ‘clumping at zero’, or be bimodal. For example, an ordinal model with a single binary covariate representing females (reference category) and males has intercepts that pertain to females and are easily translated by themselves into a cumulative probability distribution of Y for females.
- Because of the previous point, semiparametric models allow one to estimate effect ratios, exceedance probabilities $\Pr(Y \geq y|X)$, differences in exceedance probabilities (e.g., absolute risk reduction due to treatment as a function of X), quantiles of $Y|X$, and if Y is interval-scaled, the mean of $Y|X$.
- Semiparametric models do assume that there is a systematic shift in the Y distribution as one moves from one covariate setting to another. In the single binary covariate female-male example, the regression coefficient for being male completely defines how the distribution of Y for females is shifted to get the distribution for males.
 - Cox proportional hazards (PH) model: the shift in the survival distribution (one minus the cumulative distribution P_1 of Y for females) is exponentiated by a hazard

¹One exception is the continuation ratio ordinal logistic model that is covered in Chapters 13 and 14 of [Regression Modeling Strategies](#). That model is a discrete hazard-based proportional hazards model and applies primarily to a discrete dependent variable Y .

ratio r (anti-log of the regression coefficient) to obtain one minus the cumulative distribution P_2 for males, i.e. $1 - P_2 = (1 - P_1)^r$. The two distributions are assumed to be parallel on the log-log scale.

- Proportional odds (PO) ordinal logistic regression model: the cumulative distribution for females is shifted by an odds ratio r (anti-log of the sex regression coefficient) to get the cumulative distribution P_2 for males in this way: $\frac{P_2}{1-P_2} = r \times \frac{P_1}{1-P_1}$. The two distributions are assumed to be parallel on the logit (logs odds; $\log(\frac{P}{1-P})$) scale.
- Contrast semiparametric model assumptions with those of a parametric model. For the Gaussian linear model, parallelism of the normal inverse of two cumulative distribution functions corresponds to an equal variance assumption, and the curves need to be straight (i.e., a Gaussian distribution holds), not just parallel. In addition, one must be confident that Y has been properly transformed, an irrelevant requirement for semiparametric models.
- Regression coefficients and intercepts do not change if Y is transformed. Neither do predicted cumulative probabilities or quantiles (e.g. the first quartile of transformed Y is the transformation of the first quartile of untransformed Y). Only predicted means are not preserved under Y transformation.
- Since only the rank order of Y is used, the models are robust to outliers in Y (but not in X).
- The models work equally well for discrete as for continuous Y . One can have more parameters in the model than the number of observations due to the large number of intercepts when Y is continuous and has few ties. Since these intercepts are forced to be in order (in the cumulative probability model family we are dealing with), the effective number of parameters estimated is much smaller. Note that the only efficient full likelihood software implementation at present (besides SAS JMP) for continuous Y in large datasets is the `orm` function in the R `rms` package, which can efficiently handle more than 6000 intercepts (one less than the number of distinct Y values).
- Ordinal models allow one to analyze a continuous response variable that is overridden by clinical events. For example, in analyzing renal function measured by serum creatinine one could override the creatinine measurement with any arbitrary number higher than the highest observed creatinine when a patient requires dialysis. This clumping at the rightmost value of the distribution presents no problems to an ordinal model.
- When one uses ordinary regression to analyze a response that is ordinal but not interval scaled, [bad things happen](#).

The most popular semiparametric models are the Cox PH model and the PO ordinal logistic model. Cox developed a partial likelihood method so that the intercepts could be estimated separately and only the regression coefficients need to be optimized. Other semiparametric models use a full likelihood approach that estimates intercepts simultaneously with regression coefficients.

2 Particular Models

Some special cases of semiparametric models are as follows.

- The log-rank test is a special case of the Cox PH model, and so it must assume proportional hazards.
- The Wilcoxon-Mann-Whitney two-sample rank-sum test and the Kruskal-Wallis test are special cases of the proportional odds model².

In randomized clinical trials, researchers often seek to avoid making a PO assumption by dichotomizing ordinal Y . The costs of this simplification are having to count events of vastly different severities as if they were the same (e.g., hospitalization same as death) and greatly reduced power/increased sample size to achieve the same power as a PO analysis. In general, designers do not weigh costs of oversimplification but only see costs of model assumptions. Dichotomization of Y involves a severe *data assumption*. One of the blog articles discusses the ramifications of the PO assumption in detail.

2.1 Example Model

Consider a discrete case with response variable $Y = 0, 1, 2, 3$ representing pain levels of none, mild, moderate, severe. Let the covariates X represent an indicator variable for sex (0=female, 1=male) and treatment (0=control, 1=active). A PO model for Y could be

$$\Pr(Y \geq y|X) = \text{expit}(\alpha_y + \beta_1[\text{male}] + \beta_2[\text{active}])$$

where $\text{expit}(z) = \frac{1}{1+\exp(-z)}$, $\alpha_1 = 1, \alpha_2 = 0, \alpha_3 = -1, \beta_1 = -0.5, \beta_2 = -0.4$. The male:female OR for $Y \geq y$ for any y is $\exp(-0.5) = 0.61$ and the active:control OR is $\exp(-0.4) = 0.67$.

Consider probabilities of outcomes for a male on active treatment. The β part of the model is -0.9. The probabilities of outcomes of level y or worse are as follows:

y	Meaning	log odds($Y \geq y$)	$\Pr(Y \geq y)$
1	any pain	0.1	0.52
2	moderate or severe	-0.9	0.29
3	severe	-1.9	0.13

The probability that the pain level will be moderate is $0.29 - 0.13 = 0.16$. The probability of

²Multiple explanations for the Wilcoxon test assuming proportional odds are given [here](#)

being pain free is $1 - 0.52 = 0.48$.

A model for continuous Y would look the same, just have many more α s.

3 Tutorials and Course Material

For a gentle introduction see Section 7.6 in Biostatistics for Biomedical Research, below.

Regression Modeling Strategies [Text](#) and its [Online Course Notes](#)

- Chapter 13: Ordinal logistic regression
- Chapter 14: Detailed case study for a discrete ordinal Y
- Chapter 15: Ordinal regression for continuous Y with detailed case study

Biostatistics in Biomedical Research [PDF](#) and [Chapters in HTML](#)

Topic	Section
Ordinal outcomes in clinical trials	3.6 and 5.12.5
Proportional odds model	7.6
Power calculations tailored to the proportional odds model	7.8.3
Bayesian logistic model	6.10.3

4 Arguments In Favor of Wider Use of the PO Model

For reasons that can only be explained by unfamiliarity, many reviewers question the reliance on ordinal model assumptions even when they do not insist on verification of equal variance, normality, or proportional hazards assumptions in other settings. Some arguments to assist in interactions with such reviewers are the following.

- All statistical methods have assumptions
 - t -test: normality, equal variance; QQ plots: two parallel **straight lines** of normal inverse cumulative distributions
 - PO model: two parallel **any shape** curves of logit of cumulative distributions

- Assumptions of PO model far less stringent than the t -test (PO uses only rank ordering of Y)
- Assumptions
 - The PO assumption is a **model assumption** that may or may not hold
 - Splitting an ordinal outcome at a cutpoint is a **data assumption** that is **known** not to hold
 - * assuming e.g. hospitalization is the same as death
 - Interest is often in whether patients improve, i.e., whether there is a shift in their outcome distribution from being on different treatments. Interest is often not confined to whether a single threshold was achieved.
 - Just as with the proportional hazards assumption, PO analysis can provide meaningful summary ORs even when PO is violated, unless the violation is in the form of a major reversal in the direction of the treatment effect for certain cutoffs (where the sample sizes support cutoff-specific estimation)
- Examining treatment effects over all cutoffs of the outcome for examining the PO assumption is not reliable (wide confidence intervals for ORs)
 - Designing a study to be able to test/estimate a specific cutoff's OR requires much larger sample size than using the whole ordinal spectrum
- PO = Wilcoxon test when there is no covariate adjustment → **The PO model is an extension of the Wilcoxon test handling covariates**
 - Wilcoxon test is an accepted way to assess whether one treatment has higher patient responses than other, when the response is continuous or ordinal
 - Scaling the Wilcoxon statistic to $[0,1]$ yields a concordance probability c (which is the Mann-Whitney U statistic)
 - * Probability that a randomly chosen patient on treatment B has a higher Y value than a randomly chosen patient on treatment A
 - Over a wide variety of simulated clinical trials the R^2 between the PO model log OR and $\text{logit}(c)$ is 0.996
 - Mean absolute error in computing the Wilcoxon statistic (scaled to $[0,1]$) from the OR is 0.008
 - Simple conversion formula $c = \frac{\text{OR}^{0.65}}{1+\text{OR}^{0.65}}$
 - The numerator of the score χ^2 statistic from the PO model is identical to the Wilcoxon statistic
 - For detailed simulations with all R code provided see <https://fharrell.com/post/po-wilcoxon>
- Interpretation of PO model results

- Treatment OR is interpreted just as OR from a binary outcome, just need to pool outcomes
- E.g. OR for $Y \geq 3$ vs. $Y < 3$ in place of OR for $Y = 1$ vs. $Y = 0$
- Side-by-side stacked bar charts
 - * proportion of patients at each level of Y for each treatment
 - * same but from the model for a specific covariate setting

5 Testing the PO Assumption

Peterson and Harrell (1990) *Applied Stat* **39**:205-217 showed that the score test for proportional hazards is anti-conservative, i.e., can have highly inflated α from p -values being too small³. As Peterson and Harrell state on p. 215:

The simulations reveal that both tests often give blatantly erroneous results when the cross-tabulation table for the response variable by an explanatory variable contains empty cells at an inner value of Y , i.e., $1 < Y < k$. Less blatant, but still suspicious, results are occasionally obtained if the table suffers from a general sparseness of cell sizes. The score test also suffers if the number of observations at one of the levels of Y is small relative to the total sample size.

Peterson and Harrell did not study operating characteristics of the likelihood ratio (LR) χ^2 test for proportional odds, but in general LR tests have better performance than Wald and score tests⁴. See [this](#), which emphasizes the importance of assessing the (minimal) impact of the PO assumption rather than testing it, for details

6 Relaxing the PO Assumption

Peterson and Harrell (1990) developed the partial PO model to relax the PO (equal slopes / parallelism) assumption for one or more predictors. See **Software** below for implementations.

7 Blog Articles on [Statistical Thinking](#)

- [Violation of proportional odds is not fatal](#)
- [If you like the Wilcoxon test you must like the proportional odds model](#)
- [Assessing the proportional odds assumption and its impact](#)
- [Information gain from using ordinal instead of binary outcomes](#)

³This was pointed out to the author of SAS PROC LOGISTIC in 1990, and the author elected to ignore this finding in implementing the very problematic score test in the procedure.

⁴See [this paper](#) by Andrew Thomas who studied the likelihood ratio test for PO.

- [Equivalence of Wilcoxon test and proportional odds model](#)

8 Other Resources

- R rmsb package [Bayesian proportional odds model with random effects](#) and implementing the partial proportional odds model to allow pre-specified departures from the proportional odds assumption
- Nathan James' R package for the [Bayesian proportional odds model](#)
- [Stan code](#) from Ben Goodrich (Columbia University) for the PO model and [Ben's notes](#)
- [Inner workings of Stan ordinal models](#) and [Github](#)
- [Simulating ordinal outcome data](#) by Keith Goldfeld
- [Risk prediction models for discrete ordinal outcomes: Calibration and the impact of the proportional odds assumption](#) by M Edlinger, M van Smeeden, H Alber, M Wanitschek, B Van Calster

9 Longitudinal Ordinal Models

- [Resources home page](#)
- [References](#) related to ordinal longitudinal and Markov models
- [Papers](#) about statistical modeling of longitudinal ordinal responses

10 Clinical Trial Design Resources

- [Power and sample size calculation](#), Section 7.8
- [Paper](#) exemplifying game playing in choice of RCT endpoint and how things go seriously wrong with time to symptom resolution as an endpoint (authors had to ignore emergency department visits)
- [Formulating outcomes in clinical trials](#)

11 Contrasting Proportional Odds Model Analysis with Dichotomized Outcome Analysis

- [PO Model vs. Binary Logistic Model Map](#)
- [Power, precision, and arbitrariness problems](#) from **not** using an ordinal model

12 Software

R

- [rms package](#)
 - [lrm function](#) for discrete Y or continuous Y having not more than a couple of hundred distinct levels, only implements the logit link
 - [orm function](#) is intended for continuous Y and efficiently handles up to several thousand distinct values; multiple link functions are handled
- [VGAM package](#) for discrete Y implements a huge number of link functions and allows for general relaxation of proportional odds and related assumptions, allowing one to fit the partial proportional odds model
- [rmsb package](#)
 - [blrm function](#) implements Bayesian proportional odds, partial PO, and constrained partial PO models with optional random intercepts for Y with up to perhaps 200 distinct values (execution time is linear in the number of distinct Y)
- Hmisc package [popower](#) and [posamsize](#) functions for power and sample size calculations for unadjusted PO comparisons and the Wilcoxon test

Other

All major statistical software packages have a proportional odds model. The partial PO model is probably only implemented in R, and only SAS JMP has an efficient implementation like `rms::orm` for continuous Y.

13 General References

- <https://hbiostat.org/bib/ordinal.html> which includes tutorials such as the one by Susan Scott et al.