Received: 30 April 2018 | Revised: 14 December 2019 | Accepted: 6 January 2020
DOI: 10.1002/sim.8479

RESEARCH ARTICLE

Statistics in Medicine WILEY

Ordinal outcomes: A cumulative probability model with the log link and an assumption of proportionality

Gurbakhshash Singh[1] | **Gordon Hilton Fick**[2]

[1]Department of Mathematical Sciences, Central Connecticut State University, New Britain, Connecticut

[2]Department of Community Health Sciences, University of Calgary, Alberta, Canada

Correspondence
Gurbakhshash Singh, Department of Mathematical Sciences, Central Connecticut State University, 1615 Stanley Street, PO Box 4010, New Britain, CT 06053.
Email: gsingh@ccsu.edu

Summary

We present here a study of ordinal outcomes with a cumulative probability model. In particular, we consider the log link with the assumption of proportionality. The logit link is currently the most widely used link with ordinal outcomes in the health research literature. With the logit link, one obtains regression coefficients that are functions of odds. When the log link is used, one obtains regression coefficients that are functions of probabilities. While odds might be preferred with certain studies that are retrospective, many health researchers may prefer to have direct statements about probabilities. There are two classes of models with the log link analogous to those with the logit link. We will call these two classes the Proportional Probability Model (PPM) and the Log Cumulative Probability Model (LCPM). These models introduce a challenge not seen with the logit link models. The log link models have constraints on the parameter space. We must insist that the maximum likelihood estimate (MLE) satisfy these constraints. We present conditions for the uniqueness of the MLE and we present other features of the MLE. Several examples and several closed form expressions for the MLE are presented. While this paper is primarily about the PPM, our R package `lcpm` contains functions to fit both the PPM and the LCPM.

KEYWORDS

proportional probability model, maximum likelihood estimation, uniqueness, constrained maximization

1 | INTRODUCTION

Models for an ordinal outcome were presented by Aitchison and Silvey[1] and later by McCullagh.[2] They studied Generalized Linear Models with the probit link and the logit link, respectively. Recently, Blizzard et al[3] have used the log link. Blizzard suggests the use of the log link for many different models. In these models, we will show that the use of the log link is now realistically available for health researchers in many settings. We will see that the log link introduces linear constraints upon the parameter space.

Let the outcome variable y be a categorical response with J ordered categories. With the covariate vector $\boldsymbol{x}_i'$, the Log cumulative probability model (LCPM) is $\log[P(y \leq j|\boldsymbol{x}_i)] = \alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta}_j$ for cuts $j = 1, \ldots, (J-1)$ and subjects $i = 1, \ldots, n$.

Abbreviations: PPM, Proportional Probability Model; LCPM, Log Cumulative Probability Model; MLE, Maximum Likelihood Estimate; CFE, Closed Form Expression

There are parameters α_j and $\boldsymbol{\beta}_j$ for each cut j. The proportional probability model (PPM) is an LCPM which assumes the same $\boldsymbol{\beta}_j = \boldsymbol{\beta}$ for all cuts j. The PPM assumes that $P(y \leq j)$ and $P(y \leq j')$ are proportional. The constant of proportionality is $e^{\alpha_j - \alpha_{j'}}$ for $j \neq j'$. The linear inequality constraints in the PPM are

1. from the log link: $\alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta} \leq 0$ for $i = 1, \dots, n$ and $j = 1, \dots, (J-1)$ which ensures $0 \leq P(y \leq j|\boldsymbol{x}_i) \leq 1$ and
2. from the cumulative probability: $\alpha_j \leq \alpha_{j+1} \leq 0$ which ensures the cumulative probability is monotonic increasing.

Computation of the maximum likelihood estimate (MLE) for the PPM requires optimization methods that can manage the linear inequality constraints. Blizzard et al[3] developed methods using `PROC CATMOD` in SAS but their proposed MLE might not satisfy the constraints. Prior to the models formulated by Blizzard, Williams[4] developed `oglm` in Stata for ordinal regression models. This methodology permitted the use of a log link, but we regard it as preliminary to our work. Indeed, Williams noted "that link(log) is considered experimental and possibly wrong"[5] in a help file. To our knowledge, this early work had not been implemented in R.[6] The `VGAM`[7] package using function `vglm` does not manage the inequality constraints on the parameter space. Without the constraints, incorrect parameter estimates can result in clearly incorrect probability estimates that are larger than one, see Appendix A.2 for an example. We have introduced the R package `lcpm`[8] which contains preliminary function `ppm` to determine the MLE for a PPM with linear inequality constraints on the parameter space.

There is a lengthy history[9-29] of proposed MLE and non-MLE methods for estimating functions of probabilities (like ratios of probabilities, risk ratios, and other forms). Maximum likelihood estimation with a Log Binomial Model[9] (LBM) subject to constraints $\boldsymbol{x}_i'\boldsymbol{\beta} \leq 0, \forall i$ is a similar problem to the MLE determination with the PPM. Recent developments in the LBM,[25,29] have used the function `constrOptim` in R. This function uses an adaptive barrier algorithm of Lange[30,31] which manages the linear inequality constraints of the LBM. We use `constrOptim` in our package `lcpm` for determining the MLE of the PPM with the linear inequality constraints given above.

This article provides results for the MLE with the PPM . In Sections 3.3 and A.1, the conditions for the uniqueness of the MLE are developed. Section 3.4 develops closed form expressions (CFEs) for the MLE in certain settings. These will be used in Section 4 to assess the use of the function `ppm` in determining the MLE. A brief discussion of conditions for an MLE on a boundary is presented in Section 5.

2 | NOTATION

In this section, we introduce some notation that will be used throughout the article. Let j denote the ordinal (ordered) categories of an outcome variable y. Let $\boldsymbol{y}_i' = (y_{i1}, y_{i2}, \dots, y_{iJ})$ have components y_{ij} which denote the indicator for observation unit i and category j, where $i = 1, \dots, n$. For the categories, let the cumulative probability be $P(y \leq j|\boldsymbol{x}_i) = \pi_1(\boldsymbol{x}_i) + \dots + \pi_j(\boldsymbol{x}_i)$ where $P(y = j|\boldsymbol{x}_i) = \pi_j(\boldsymbol{x}_i)$, $\boldsymbol{x}_i'$ is ith row of covariates of the $n \times p$ covariate matrix $\boldsymbol{X}$ and $\sum_j \pi_j(\boldsymbol{x}_i) = 1$. The $\boldsymbol{y}_i'$ are then each categorical distributions with probabilities $(\pi_1(\boldsymbol{x}_i), \dots, \pi_J(\boldsymbol{x}_i))$ where $\{\pi_1(\boldsymbol{x}_i) = P(y = 1|\boldsymbol{x}_i), \pi_j(\boldsymbol{x}_i) = P(y \leq j|\boldsymbol{x}_i) - P(y \leq j-1|\boldsymbol{x}_i)$ for $j = 2, \dots J-1$ and $\pi_J(\boldsymbol{x}_i) = 1 - P(y \leq J-1|\boldsymbol{x}_i)\}$. The $\boldsymbol{y}_i'$ are assumed to be statistically independent. We then have:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \quad \text{and corresponding outcomes} \quad \begin{bmatrix} \boldsymbol{y}_1' \\ \boldsymbol{y}_2' \\ \vdots \\ \boldsymbol{y}_n' \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1J} \\ y_{21} & y_{22} & \dots & y_{2J} \\ \dots & \dots & \dots & \dots \\ y_{n1} & \dots & \dots & y_{nJ} \end{bmatrix}.$$

We will now see that we can group by distinct covariate patterns. Let us define a second set of matrices $\boldsymbol{C}$ and $\boldsymbol{N}$. $\boldsymbol{C}$ contains the distinct covariate rows of $\boldsymbol{X}$. We suppose that $\boldsymbol{C}$ is full column rank with $m \leq n$ different groups or covariate sets. For each row of $\boldsymbol{C}$, let n_{kj} denote the count of the number of outcomes for group k and category j. Let $\boldsymbol{N}$ denote a matrix of the (n_{kj}) counts of the number of outcomes for each group k and ordinal categories j. Also let $\boldsymbol{N}^{(j)}$ be jth column of $\boldsymbol{N}$ and let $\boldsymbol{N}^{(-j)}$ be $\boldsymbol{N}$ without the jth column. The matrices $\boldsymbol{C}$ and $\boldsymbol{N}$ are

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \dots & \dots & \dots & \dots \\ c_{m1} & \dots & \dots & c_{mp} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1' \\ \boldsymbol{c}_2' \\ \vdots \\ \boldsymbol{c}_m' \end{bmatrix} \quad \text{and corresponding counts} \quad \boldsymbol{N} = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1J} \\ n_{21} & n_{22} & \dots & n_{2J} \\ \dots & \dots & \dots & \dots \\ n_{m1} & \dots & \dots & n_{mJ} \end{bmatrix} = \begin{bmatrix} \boldsymbol{n}_1' \\ \boldsymbol{n}_2' \\ \vdots \\ \boldsymbol{n}_m' \end{bmatrix}.$$

Denote the sum of column j of $\boldsymbol{N}$ as $n^{(j)} = \sum_{k=1}^{m} n_{kj}$ and denote the sum of row k of $\boldsymbol{N}$ as $n_{(k)} = \sum_{j=1}^{J} n_{kj}$. Note that $n^{(j)} > 0$ and $n_{(k)} > 0$, otherwise the respective information for the column (category) or row (group) can be removed from the matrix $\boldsymbol{N}$. We can also note that $\boldsymbol{n}'_{\boldsymbol{k}}$ is multinomial with the number of trials $n_{(k)}$ and probabilities $(\pi_1(\boldsymbol{c}_{\boldsymbol{k}}), \dots, \pi_J(\boldsymbol{c}_{\boldsymbol{k}})))$ where the $\boldsymbol{n}'_{\boldsymbol{k}}$ are statistically independent.

3 | THE PROPORTIONAL PROBABILITY MODEL

If we now assume that β is the same for each cut (j), we will call this model the PPM:

$$\log[P(y \leq j|\boldsymbol{x}_{\boldsymbol{i}})] = \alpha_j + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta}.$$

As is the case with the proportional odds model, the β are said to be assumed common to the cuts (j) and the α_j are assumed common to the $\boldsymbol{\beta}$. The estimation involves $(J-1)$ components of $\boldsymbol{\alpha}$ and p components of $\boldsymbol{\beta}$. We note that there are restrictions on the parameter space Ω_{PPM}. The log of a probability is negative and so $\alpha_j\mathbf{1} + \boldsymbol{X}\boldsymbol{\beta} \leq \mathbf{0}$ for all j. The cumulative probability must increase with each successive cut and so $\alpha_{j+1} \geq \alpha_j$, $\forall j$. The parameter space has constraints: $\Omega_{PPM} = \{(\boldsymbol{\alpha},\boldsymbol{\beta})|\alpha_j + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta} \leq 0, \alpha_{j+1} \geq \alpha_j, j = 1, \dots, (J-1), \forall i\}$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{J-1})'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$. In Section 5 we provide consideration to the boundary of Ω_{PPM} which is $\{(\boldsymbol{\alpha}, \boldsymbol{\beta})|\alpha_j + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta} = 0\} \cup \{\boldsymbol{\alpha}|\alpha_{j+1} = \alpha_j\}$.

The likelihood function can be expressed as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = c\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\pi_j(\boldsymbol{x}_{\boldsymbol{i}})^{y_{ij}}\right].$$

The log-likelihood function is

$$\begin{aligned}\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = &\sum_{i=1}^{n} y_{i1}\log\left(\exp(\alpha_1 + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta})\right) + \sum_{i=1}^{n}\sum_{j=2}^{J-1} y_{ij}\log\left(\exp(\alpha_j + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta}) - \exp(\alpha_{j-1} + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta})\right)\\ &+ \sum_{i=1}^{n} y_{iJ}\log\left(1 - \exp(\alpha_{J-1} + \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta})\right) + a.\end{aligned}$$

The log-likelihood can be rewritten using the groups introduced in the previous section:

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{k=1}^{m} n_{k1}\left(\alpha_1 + \boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta}\right) + \sum_{k=1}^{m}\sum_{j=2}^{J-1} n_{kj}\left(\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta} + \log(\exp(\alpha_j) - \exp(\alpha_{j-1}))\right) + \sum_{k=1}^{m} n_{kJ}\log\left(1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})\right) + a,$$

3.1 | lcpm R package

Determining the MLE requires finding $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ that maximizes $\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})$, subject to the constraints on the parameter space Ω_{PPM}. Our R package `lcpm` contains a function `ppm` which determines the MLE for the PPM using the R function `constrOptim`.

`constrOptim` is a part of the `stats` package in R. It provides an adaptive barrier optimization algorithm[30] for the minimization of a function subject to equality and inequality constraints. This adaptive barrier approach turns out to be precisely what we need to determine the MLE. To briefly introduce these details, we note that this minimization can be accomplished by constructing a surrogate function $\ell^*(\boldsymbol{\alpha}, \boldsymbol{\beta})$, which consists of a barrier function $b(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and a positive tuning parameter μ. The surrogate function $\ell^*(\boldsymbol{\alpha}, \boldsymbol{\beta}) = -\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \mu b(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is then minimized over decreasing values of μ (that is as $\mu \to 0$, $\ell^* \to -\ell$). The algorithm addresses the minimization problem of a convex function subject to the concave inequality constraints. The PPM inequality constraints in `constrOptim` must be expressed as $-\alpha_j - \boldsymbol{x}'_{\boldsymbol{i}}\boldsymbol{\beta} \geq 0$ and $\alpha_{j+1} - \alpha_j \geq 0$. This algorithm requires the specification of inequality constraints, the selection of feasible starting values, and the monitoring of circumstances when the MLE is on or near the boundary.

3.2 Score and Hessian

The score function $\boldsymbol{S}(\boldsymbol{\alpha},\boldsymbol{\beta}) = (\boldsymbol{S}'_\alpha, \boldsymbol{S}'_\beta)'$ is a $(J-1+p)\times 1$ vector consisting of $(J-1)\times 1$ vector $\boldsymbol{S}'_\alpha$ of derivatives of the log-likelihood with respect to α_j for $j = 1, \dots, (J-1)$ and $p \times 1$ vector $\boldsymbol{S}'_\beta$ of derivatives of the log-likelihood with respect to β_l for $l = 1, \dots, p$. The elements of $\boldsymbol{S}_\beta$ are given by:

$$\frac{\partial \ell(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \beta_l} = \sum_{k=1}^{m}\sum_{j=1}^{J-1} n_{kj}c_{kl} - \sum_{k=1}^{m} n_{kJ}c_{kl}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})}{1-\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})},$$

which is defined for $\{(\alpha_{J-1},\boldsymbol{\beta})|\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta}<0\}$. The elements of $\boldsymbol{S}_\alpha$ are given by:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \alpha_1} &= n^{(1)} - n^{(2)}\frac{\exp(\alpha_1)}{\exp(\alpha_2)-\exp(\alpha_1)},\\ \frac{\partial \ell(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \alpha_{J-1}} &= n^{(J-1)}\frac{\exp(\alpha_{J-1})}{\exp(\alpha_{J-1})-\exp(\alpha_{J-2})} - \sum_{k=1}^{m} n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})}{1-\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})},\end{aligned} \tag{1}$$

and for $1 < j < J-1$

$$\frac{\partial \ell(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \alpha_j} = -n^{(j+1)}\frac{\exp(\alpha_j)}{\exp(\alpha_{j+1})-\exp(\alpha_j)} + n^{(j)}\frac{\exp(\alpha_j)}{\exp(\alpha_j)-\exp(\alpha_{j-1})}, \tag{2}$$

which is defined for $\{(\boldsymbol{\alpha},\boldsymbol{\beta})| -\infty < \alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta} < 0, \alpha_{j+1} > \alpha_j\}$. As noted by Liu et al,[32] the Hessian ($\boldsymbol{H}$) will be a patterned matrix. The observed Fisher Information matrix ($-\boldsymbol{H}$) is used in the estimation of the variance of the parameters.[33]

$$\boldsymbol{H} = \left[\begin{array}{c|c}\boldsymbol{H}_{\alpha\alpha} & \boldsymbol{H}_{\alpha\beta}\\ \hline \boldsymbol{H}'_{\alpha\beta} & \boldsymbol{H}_{\beta\beta}\end{array}\right].$$

In assessing the terms of the Hessian, denote

$$\delta_k = n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}'_{\boldsymbol{k}}\boldsymbol{\beta}))^2} \quad \text{and} \quad \gamma_{k,j+1} = n_{k(j+1)}\frac{\exp(\alpha_{j+1}+\alpha_j)}{(\exp(\alpha_{j+1})-\exp(\alpha_j))^2}.$$

$\boldsymbol{H}_{\alpha\alpha}$ is a $(J-1)\times(J-1)$ triband matrix of the associated second partial derivatives with respect to $\boldsymbol{\alpha}$ which has components

$$A_{st} = \frac{\partial^2 \ell}{\partial \alpha_s \partial \alpha_t} \quad \text{and is} \quad \boldsymbol{H}_{\alpha\alpha} = \begin{bmatrix} A_{11} & A_{12} & 0 & 0 & \dots & 0 & 0 & 0\\ A_{12} & A_{22} & A_{23} & 0 & \dots & 0 & 0 & 0\\ 0 & A_{23} & A_{33} & A_{34} & \dots & 0 & 0 & 0\\ 0 & 0 & A_{34} & A_{44} & \dots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \dots & A_{J-3,J-3} & A_{J-2,J-3} & 0\\ 0 & 0 & 0 & 0 & \dots & A_{J-2,J-3} & A_{J-2,J-2} & A_{J-2,J-1}\\ 0 & 0 & 0 & 0 & \dots & 0 & A_{J-2,J-1} & A_{J-1,J-1}\end{bmatrix}.$$

The second derivatives are as follows: $A_{11} = -\sum_{k=1}^{m}\gamma_{k2}$, $A_{J-1,J-1} = \sum_{k=1}^{m}(-\delta_k - \gamma_{k(J-1)})$, $A_{12} = \sum_{k=1}^{m}\gamma_{k2}$, $A_{J-2,J-1} = \sum_{k=1}^{m}\gamma_{k(J-1)}$, and for $1 < j < J-1$: $A_{jj} = \sum_{k=1}^{m}(-\gamma_{k(j+1)} - \gamma_{kj})$ and $A_{j+1,j} = \sum_{k=1}^{m}\gamma_{k(j+1)}$.

$\boldsymbol{H}_{\alpha\beta}$ is a $(J-1)\times p$ matrix of zeros with one row consisting of the mixed partial derivatives with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which has components

$$M_l = \frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \alpha_{J-1}\partial \beta_l} = -\sum_{k=1}^{m} c_{kl}\delta_k \quad \text{and is} \quad \boldsymbol{H}_{\alpha\beta} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0\\ M_1 & M_2 & M_3 & M_4 & \dots & M_{p-2} & M_{p-1} & M_p\end{bmatrix}.$$

$\boldsymbol{H}_{\beta\beta}$ is a $p\times p$ matrix of the second partial derivatives with respect to $\boldsymbol{\beta}$ which has components

$$B_{ol} = \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_o \partial \beta_l} = -\sum_{k=1}^{m} c_{kl} c_{ko} \delta_k \text{ and is } \boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1(p-1)} & B_{1p} \\ B_{21} & B_{22} & \dots & B_{2(p-1)} & B_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ B_{(p-1)1} & B_{(p-1)2} & \dots & B_{(p-1)(p-1)} & B_{(p-1)p} \\ B_{p1} & B_{p2} & \dots & B_{p(p-1)} & B_{pp} \end{bmatrix}.$$

3.3 | Uniqueness of the MLE for the PPM

An MLE will be unique if there is only one maximum of the log-likelihood function. For generalized linear models, Wedderburn[34] used strict concavity[35] of the log-likelihood function to determine conditions for which there is at most one maximum of the log-likelihood function. This section contains the conditions for the uniqueness of the MLE for a PPM.

Recall the partitioning of the Hessian from Section 3.2. Let $\boldsymbol{C}^* = [\boldsymbol{1} \vdots \boldsymbol{C}]$ be a $m \times (p+1)$ matrix which attaches the $m \times 1$ vector of ones to the $m \times p$ matrix $\boldsymbol{C}$. Also let $I = \{k | n_{kJ} = 0\}$ be an index set and $\boldsymbol{C}^*_{-\boldsymbol{I}}$ be a $(m - |I|) \times (p+1)$ matrix of $\boldsymbol{C}^*$ without rows given by I.

Theorem 1. *For the PPM and for* $\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) | -\infty < \alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta} < 0, \alpha_{j+1} > \alpha_j\}$:

1. *If* $\boldsymbol{C}^*_{-\boldsymbol{I}}$ *is full column rank then* $\boldsymbol{H}$ *is negative definite and the log-likelihood is strictly concave.*
2. *If* $\boldsymbol{C}^*_{-\boldsymbol{I}}$ *is not full column rank then* $\boldsymbol{H}$ *is negative semi-definite and the log-likelihood is concave.*

The details and the proof of Theorem 1 is presented in Appendix A.1. Theorem 1 shows that the log-likelihood for the PPM is concave with restrictions for strict concavity based on counts in $\boldsymbol{N}$ and if $\boldsymbol{C}^*_{-\boldsymbol{I}}$ is full column rank. These conditions are assessed using tools in R.

3.4 | Closed form expressions

We now offer some CFEs for the MLE. These expressions enable the comparison of MLEs determined by numerical methods against the CFE. Let J be the total possible ordinal values for response y and let k be the total number of possible covariate sets and that $n^{(j)} > 0, \forall j$. The score function from Section 3.2 is used to determine critical points.

Since $n^{(1)} > 0$ and $n^{(2)} > 0$, setting $\frac{\partial \ell}{\partial \alpha_1} = 0$ from Equation (1) and rearranging gives: $\exp(\alpha_2 - \alpha_1) = (\frac{n^{(2)}}{n^{(1)}} + 1)$ and for $1 < j < J-1$, since $n^{(j+1)} > 0$ and $n^{(j)} > 0$, setting $\frac{\partial \ell}{\partial \alpha_j} = 0$ from Equation (2) and rearranging gives: $\exp(\alpha_{j+1} - \alpha_j) = \left(\frac{\sum_{q=1}^{j+1} n^{(q)}}{\sum_{q=1}^{j} n^{(q)}} \right)$ and $\exp(\alpha_{J-1} - \alpha_{J-2}) = \left(\frac{\sum_{q=1}^{J-1} n^{(q)}}{\sum_{q=1}^{J-2} n^{(q)}} \right)$. Apply these expressions in the following derivative:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_{J-1}} &= n^{(J-1)} \frac{\exp(\alpha_{J-1})}{\exp(\alpha_{J-1}) - \exp(\alpha_{J-2})} - \sum_{k=1}^{m} n_{kJ} \frac{\exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})} \\ &= \sum_{q=1}^{J-1} n^{(q)} - \sum_{k=1}^{m} n_{kJ} \frac{\exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}. \end{aligned} \tag{3}$$

Note that $\sum_{q=1}^{J-1} n^{(q)}$ can be rewritten as $\sum_{k=1}^{m} \sum_{j=1}^{J-1} n_{kj}$. Now the remaining derivatives with respect to α_{J-1} and β_l are written as

$$\begin{aligned} \begin{bmatrix} \frac{\partial \ell}{\partial \alpha_{J-1}} \\ \frac{\partial \ell}{\partial \beta_1} \\ \vdots \\ \frac{\partial \ell}{\partial \beta_p} \end{bmatrix} &= \begin{bmatrix} \sum_{k=1}^{m} \sum_{j=1}^{J-1} n_{kj} - \sum_{k=1}^{m} n_{kJ} \frac{\exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})} \\ \sum_{k=1}^{m} \sum_{j=1}^{J-1} n_{kj} c_{k1} - \sum_{k=1}^{m} n_{kJ} c_{k1} \frac{\exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})} \\ \vdots \\ \sum_{k=1}^{m} \sum_{j=1}^{J-1} n_{kj} c_{kp} - \sum_{k=1}^{m} n_{kJ} c_{kp} \frac{\exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}'_{\boldsymbol{k}} \boldsymbol{\beta})} \end{bmatrix} \\ &= \boldsymbol{C}^{*\prime} \boldsymbol{N}^{(-J)} \boldsymbol{1} - \boldsymbol{\Lambda}' \mathrm{diag}(\boldsymbol{N}^{(J)}_{-\boldsymbol{I}}) \boldsymbol{\Delta}_{-I} \end{aligned} \tag{4}$$

where $\boldsymbol{C}^* = [\boldsymbol{1} \vdots \boldsymbol{C}]$ denotes a ones vector attached to the $\boldsymbol{C}$ matrix, $\boldsymbol{N}^{(-J)}$ is a $m \times (J-1)$ matrix of the $\boldsymbol{N}$ matrix without the Jth column, $\mathbf{1}$ is a $(J-1) \times 1$ vector of ones, $\boldsymbol{\Lambda} = \boldsymbol{C}^*_{-\boldsymbol{I}}$ is a $(m-|I|) \times (p+1)$ matrix of $\boldsymbol{C}^*$ without the rows given by index set I, $\boldsymbol{N}^{(J)}_{-\boldsymbol{I}}$ is a $(m-|I|) \times 1$ vector of the Jth column of the $\boldsymbol{N}$ matrix without the elements indexed by I, $\mathrm{diag}(\boldsymbol{N}^{(J)}_{-\boldsymbol{I}})$ is a $(m-|I|) \times (m-|I|)$ identity matrix with diagonal $\boldsymbol{N}^{(J)}_{-\boldsymbol{I}}$ and Δ_{-I} is a $(m-|I|) \times 1$ vector of the Δ vector without the elements indexed by I where $\Delta_{m \times 1} = (\Delta_1, \dots, \Delta_m)'$ and $\Delta_k = \frac{\exp(\alpha_{J-1} + \boldsymbol{c}_k'\boldsymbol{\beta})}{1 - \exp(\alpha_{J-1} + \boldsymbol{c}_k'\boldsymbol{\beta})}$.

The critical points are obtained by solving the system: $\boldsymbol{C}^{*\prime}\boldsymbol{N}^{(-J)}\boldsymbol{1} - \boldsymbol{\Lambda}'\mathrm{diag}(\boldsymbol{N}^{(J)}_{-\boldsymbol{I}})\boldsymbol{\Delta}_{-I} = \boldsymbol{0}_{(p+1)\times 1}$. A special case is if $n_{kJ} > 0$ for all k then critical points are obtained by solving $\boldsymbol{C}^{*\prime}\left[\boldsymbol{N}^{(-J)}\boldsymbol{1} - \mathrm{diag}(\boldsymbol{N}^{(J)})\boldsymbol{\Delta}\right] = \boldsymbol{0}_{(p+1)\times 1}$.

Theorem 2. *For the system of equations given above:*

1. *if $\boldsymbol{C}^*$ is invertible and $n_{kJ} > 0$ for all k then a closed form of MLE is given by*

$$(\alpha_{J-1}, \boldsymbol{\beta}')' = (\boldsymbol{C}^*)^{-1}\log(\boldsymbol{\Delta}/(\boldsymbol{1}+\boldsymbol{\Delta})),$$

where $\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{N}^{(J)})^{-1}\left[\boldsymbol{N}^{(-J)}\mathbf{1}\right]$ and the estimates of $(\alpha_1, \alpha_2, \dots, \alpha_{J-2})$ are given by solving the following

$$\alpha_{J-2} = \alpha_{J-1} - \log\left(\frac{\sum_{l=1}^{J-1} n^{(l)}}{\sum_{l=1}^{J-2} n^{(l)}}\right) \text{ and then backward recursively using } \alpha_j = \alpha_{j+1} - \log\left(\frac{\sum_{l=1}^{j+1} n^{(l)}}{\sum_{l=1}^{j} n^{(l)}}\right).$$

2. *if $(m-|I|) = (p+1)$ and $\boldsymbol{\Lambda}$ is invertible then a closed form of MLE is given by*

$$(\alpha_{J-1}, \boldsymbol{\beta}')' = (\boldsymbol{\Lambda})^{-1}\log(\boldsymbol{\Delta}_{-\boldsymbol{I}}/(\boldsymbol{1}+\boldsymbol{\Delta}_{-\boldsymbol{I}})),$$

where $\boldsymbol{\Delta}_{-\boldsymbol{I}} = \mathrm{diag}(\boldsymbol{N}^{(J)}_{-\boldsymbol{I}})^{-1}\boldsymbol{\Lambda}'^{-1}\left[\boldsymbol{C}^{*\prime}\boldsymbol{N}^{(-J)}\mathbf{1}\right]$ and the estimates of $(\alpha_1, \alpha_2, \dots, \alpha_{J-2})$ are given by solving the following

$$\alpha_{J-2} = \alpha_{J-1} - \log\left(\frac{\sum_{l=1}^{J-1} n^{(l)}}{\sum_{l=1}^{J-2} n^{(l)}}\right) \text{ and then backward recursively using } \alpha_j = \alpha_{j+1} - \log\left(\frac{\sum_{l=1}^{j+1} n^{(l)}}{\sum_{l=1}^{j} n^{(l)}}\right).$$

Proof of Theorem 2. The proof is constructed by solving the systems, given prior to the theorem, to determine the critical points. Then apply Theorem 1 to determine that the log-likelihood is strictly concave and the critical point is a maximum. ∎

Now notice that the CFEs are based on critical points. If the expression gives a maximum that is outside the parameter space then the MLE is on a boundary of the parameter space since the log-likelihood is strictly concave.

4 | EXAMPLES

This section contains several examples. Example 4.1 demonstrates the use of CFEs from Section 3.4 compared with the MLE found by using `ppm`. Example 4.2 demonstrates the use of the CFEs in comparison with MLE found by `ppm` for datasets found in the literature. Examples 4.3 and 4.4 provide additional applications of `ppm`. Finally, Example 4.5 shows that `ppm` can determine an MLE on a boundary.

4.1 | Simulation

This example applies the CFE from Theorem 2 where $C^*_{-I} = C^*$ is invertible and $n_{kJ} > 0$ for all k. Suppose there are $J = 3$ ordinal categories and a single binary covariate $x = \{0, 1\}$ with $m = 2$ unique covariate vectors. The PPM is:

$$\log[P(y \le j|x)] = \alpha_j + \beta_1 x,$$

for cuts $j = 1, 2$. For the PPM, $\alpha_{j+1} \geq \alpha_j$ for all j and $\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) | \alpha_j + \beta_1 x \leq 0\}$ for all j. Let $\boldsymbol{N}$ denote a matrix of the number of individuals in each of the m groups with $J = 3$ ordinal values.

$$\boldsymbol{N} = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \end{bmatrix}, \text{and} \quad \boldsymbol{C}^* = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Note that $\boldsymbol{C}^*$ is full column rank and invertible. Applying Theorem 1, the log-likelihood is strictly concave with at most one MLE in the interior of the parameter space (Ω_{PPM}).

From Theorem 2, $\boldsymbol{C}^{*\prime} \left[\boldsymbol{N}^{(-J)}\boldsymbol{1} - diag(\boldsymbol{N}^{(J)})\boldsymbol{\Delta}\right] = \boldsymbol{0}$ gives the following:

$$(\alpha_{J-1}, \boldsymbol{\beta})' = (\boldsymbol{C}^*)^{-1} \log(\boldsymbol{\Delta}/(\boldsymbol{1} + \boldsymbol{\Delta})),$$

$$\begin{bmatrix} \alpha_2 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \log(\Delta_1/(1+\Delta_1)) \\ \log(\Delta_2/(1+\Delta_2)) \end{bmatrix}$$
$$= \begin{bmatrix} \log(\Delta_1/(1+\Delta_1)) \\ \log(\Delta_2/(1+\Delta_2)) - \log(\Delta_1/(1+\Delta_1)) \end{bmatrix},$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{N}^{(J)})^{-1} \left[\boldsymbol{N}^{(-J)}\boldsymbol{1}\right]$, giving $\Delta_1 = \frac{n_{11}+n_{12}}{n_{13}}$ and $\Delta_2 = \frac{n_{21}+n_{22}}{n_{23}}$. The MLE is:

$$\hat{\alpha}_2 = \log\left(\frac{n_{11}+n_{12}}{n_{11}+n_{12}+n_{13}}\right), \tag{5}$$

$$\hat{\beta}_1 = \log\left(\frac{n_{21}+n_{22}}{n_{21}+n_{22}+n_{23}}\right) - \log\left(\frac{n_{11}+n_{12}}{n_{11}+n_{12}+n_{13}}\right), \tag{6}$$

and the estimates of $(\alpha_1, \alpha_2, \dots, \alpha_{J-2})$ are given by solving the following expression using $\alpha_{J-2} = \alpha_{J-1} - \log\left(\frac{\sum_{l=1}^{J-1} n^{(l)}}{\sum_{l=1}^{J-2} n^{(l)}}\right)$ giving

$$\hat{\alpha}_1 = \log\left(\frac{n_{11}+n_{12}}{n_{11}+n_{12}+n_{13}}\right) - \log\left(\frac{n_{11}+n_{21}+n_{12}+n_{22}}{n_{11}+n_{21}}\right). \tag{7}$$

The CFEs are compared with the MLE determined by using `ppm` in R. Table 1 displays the MLE results for the closed form and `ppm`. 15 625 datasets were constructed from n_{11}, n_{12}, n_{13}, n_{21}, n_{22} and n_{23} by taking values from vector $(1, 2, 5, 10, 100)'$. For each dataset, the MLE was computed by formula and `ppm` (using default controls in `ppm`). The absolute difference in `ppm` and CFE MLE for each parameter was on average (7.2297e-07, 4.0111e-07, 6.7437e-07) with minimum absolute difference (4.2402e-11, 4.9759e-12, 5.0844e-12) and maximum absolute difference (8.6504e-06, 7.1547e-06, 1.2319e-05). This example compared the closed form MLE with the MLE from `ppm`.

4.2 | Application 1: Tonsil Dataset

In this example, we consider the Tonsil dataset in McCullagh,[2] which is given in Table 2. There are 1398 observations pertaining to the outcome of the size of tonsils: 1 - not enlarged, 2 - enlarged, and 3 - greatly enlarged. The covariate here is an indicator variable for carrier ($x = \{0, 1\}$) for noncarriers and carriers of *Streptococcus pyogenes*. Using this data, the following model is assumed: $\log[P(y \leq j|x] = \alpha_j + \beta_1 x$, where $\boldsymbol{C}|\boldsymbol{N}$ is

$$\boldsymbol{C}|\boldsymbol{N} = \begin{bmatrix} 0 & 497 & 560 & 269 \\ 1 & 19 & 29 & 24 \end{bmatrix},$$

and where $J = 3$ and $j = 1, 2$. The CFEs (5) to (7) can be applied. The MLE from CFE and `ppm` are compared in the left portion of Table 3.

TABLE 1 MLE by CFE and ppm for Example 4.1

(n_{11}, n_{12}, n_{13}) (n_{21}, n_{22}, n_{23})	MLE	CFE	ppm
(1, 1, 1)	$\hat{\alpha}_1$	−1.09861	−1.09861
(1, 1, 1)	$\hat{\alpha}_2$	−0.40547	−0.40546
	$\hat{\beta}_1$	0	−6.58e-07
(1, 1, 1)	$\hat{\alpha}_1$	−1.32176	−1.32175
(1, 2, 10)	$\hat{\alpha}_2$	−0.40547	−0.40546
	$\hat{\beta}_1$	−1.06087	−1.06087
(1, 1, 5)	$\hat{\alpha}_1$	−1.94591	−1.94591
(2, 2, 100)	$\hat{\alpha}_2$	−1.25276	−1.25276
	$\hat{\beta}_1$	−2.00533	−2.00533
(1, 1, 2)	$\hat{\alpha}_1$	−1.73460	−1.73460
(5, 10, 100)	$\hat{\alpha}_2$	−0.69315	−0.69314
	$\hat{\beta}_1$	−1.34373	−1.34374
(1, 1, 1)	$\hat{\alpha}_1$	−1.09861	−1.09861
(10, 10, 100)	$\hat{\alpha}_2$	−0.40547	−0.40547
	$\hat{\beta}_1$	−1.38629	−1.38629
(1, 1, 5)	$\hat{\alpha}_1$	−1.94591	−1.94591
(100, 100, 10)	$\hat{\alpha}_2$	−1.25276	−1.25276
	$\hat{\beta}_1$	1.20397	1.20397
(1, 5, 100)	$\hat{\alpha}_1$	−4.66344	−4.66344
(1, 5, 100)	$\hat{\alpha}_2$	−2.87168	−2.87168
	$\hat{\beta}_1$	0	2.94e-07
(100, 1, 5)	$\hat{\alpha}_1$	−0.07523	−0.07523
(10, 2, 100)	$\hat{\alpha}_2$	−0.04832	−0.04832
	$\hat{\beta}_1$	−2.18527	−2.18527
(100, 100, 100)	$\hat{\alpha}_1$	−1.09861	−1.09861
(100, 100, 100)	$\hat{\alpha}_2$	−0.40547	−0.40547
	$\hat{\beta}_1$	0	6.19e-08

TABLE 2 Tonsil Dataset[2] for Example 4.2

Carrier status	Not enlarged	enlarged	Greatly enlarged	Total
Noncarrier	497	560	269	1326
Carrier	19	29	24	72
Total	516	589	293	1398

For the model $\log[P(y \le j|x)] = \alpha_j + \beta_1 x$, the interpretation of β_1 is common to both cuts ($j = 1, 2$) with $\beta_1 = \log[P(y \le j|x = 1)/P(y \le j|x = 0)]$. $\hat{\beta}_1$ estimates the log of the ratio of cumulative probabilities for carriers relative to noncarriers. This is an estimate of the log of a probability ratio. $e^{-0.178732} = 0.83633$ is the estimate of the ratio from the PPM.

We also note that if the outcome was reversely coded: 3—not enlarged, 2—enlarged and 1—greatly enlarged, the CFEs (5)-(7) can still be applied. The model could be rewritten as $\log[1 - P(y \le j|x)] = \kappa_j + \gamma_1 x$ and $\gamma_1 = \log[(1 - P(y \le j|x = 1))/(1 - P(y \le j|x = 0))]$. The MLE from CFE and `ppm` are compared in the right portion of Table 3. For the logit and probit links, the relationship between the MLE for the forward and reversely coded outcome is a sign change. However, for the log link there is no simple relationship between the MLE for the forward and reversely coded outcome.

TABLE 3 Maximum likelihood estimate (MLE) by closed form expression (CFE) and ppm or Example 4.2

MLE	CFE	ppm	MLE	CFE	ppm
$\hat{\alpha}_1$	−0.988226	−0.988226	$\hat{\kappa}_1$	−1.571721	−1.571721
$\hat{\alpha}_2$	−0.226732	−0.226732	$\hat{\kappa}_1$	−0.469702	−0.469703
$\hat{\beta}_1$	−0.178733	−0.178732	$\hat{\gamma}_1$	0.163328	0.163327

TABLE 4 Maximum likelihood estimate (MLE) by ppm for Example 4.3

MLE	ppm	SE	MLE	ppm	SE
$\hat{\alpha}_1$	-1.53817	0.14259	$\hat{\kappa}_1$	−0.82768	0.08097
$\hat{\alpha}_2$	−0.44587	0.08829	$\hat{\kappa}_2$	−0.22867	0.05510
$\hat{\beta}_1$	0.11979	0.11098	$\hat{\gamma}_1$	−0.11682	0.07908
$\hat{\beta}_2$	−0.39158	0.12255	$\hat{\gamma}_2$	0.10010	0.06128

4.3 | Application 2: Endometrial cancer data

We now consider data collected from a study of endometrial cancer by Hill et al.[36] The study involves 288 women diagnosed with endometrial cancer. Explanatory variables include: Race (x_1: coded as 0 – white and 1 – black) and Estrogen (x_2: coded as 0 - not used and 1 - used). Ordinal outcomes are tumor grade (coded 2 - well differentiated, 1 - moderately differentiated, and 0 - poorly differentiated). Using this data, we will explore the following model: $P[y \le j|x_1, x_2] = \alpha_j + \beta_1 x_1 + \beta_2 x_2$.

$$\boldsymbol{C}|\boldsymbol{N} = \left[\begin{array}{cc|ccc} 0 & 0 & 15 & 31 & 24 \\ 0 & 1 & 16 & 41 & 79 \\ 1 & 0 & 18 & 28 & 19 \\ 1 & 1 & 4 & 5 & 6 \end{array}\right].$$

For the model $\log[P(y \le j|x_1, x_2)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2$, the interpretation of β_2 is assumed common to both cuts and assumed common to blacks and whites with $\beta_2 = \log[P(y \le j|x_2 = 1)/P(y \le j|x_2 = 0)]$. $\hat{\beta}_2$ estimates the log of the ratio of cumulative probabilities for those receiving Estrogen and those not receiving Estrogen assumed common to race and cuts. $e^{-0.39158} = 0.67599$ is the estimate of the ratio from the PPM.

We also note that if the outcome was reversely coded: 0 - well differentiated, 1 - moderately differentiated, and 2 - poorly differentiated. The model could be rewritten as $\log[1 - P(y \le j|x_1, x_2)] = \kappa_j + \gamma_1 x_1 + \gamma_2 x_2$. The MLE is presented in the right portion of Table 4. Not only do the MLEs change but the estimated standard errors are all smaller for the reversely coded outcome model.

4.4 | Application 3: Neuropsychiatric disturbance

Now consider an example from Hosmer et al[37] and the dataset Adolescent Placement data found in R package `aplor3e`.[38] The study involves 508 subjects. The explanatory variables include: centered age (x_1), the square of centered age (x_2), state custody (x_3: coded as 0 - No, 1 - Yes), race (x_4: coded as 0 - white and 1 - non-white), and emotional disturbance (x_5: coded as 0 - not severe and 1 - severe). The ordinal outcome is neuropsychiatric disturbance (coded 0 - none, 1 - mild, 2 - moderate and 3 - severe). Using this data, we will explore the following model: $P[y \le j|x_1, x_2, x_3, x_4, x_5] = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_4 x_5$. The MLE can be found in the left portion of Table 5. The MLE is identified to be on a boundary. The position on the boundary can be identified by two subjects (one male and one female) with the following covariates: $x_1 = -1.73759, x_2 = 3.01923, x_3 = 1, x_4 = 0, x_5 = 0$.

We also note that if the outcome is reversely coded: 0 - severe, 1 - moderate, 2 - mild and 3 - none and if the model is rewritten as $\log[1 - P(y \le j|x_1, x_2, x_3, x_4, x_5)] = \kappa_j + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5 + \gamma_6 x_4 x_5$ then the MLE is presented in the right portion of Table 5 and it is not on a boundary. Not only do the MLEs from the forward coding change for the reverse coded outcome but the estimated standard errors are all larger for the reversely coded outcome model. Interestingly, the P-value associated with the interaction term is .3125 in the first model but for the reverse coded outcome the P-value is .01971. As noted earlier, the results of forward and reverse coded outcomes can lead to different estimates, SEs

MLE	ppm	SE	MLE	ppm	SE
$\hat{\alpha}_1$	−0.33664	0.03853	$\hat{\kappa}_1$	−2.732709	0.196857
$\hat{\alpha}_2$	−0.12841	0.03078	$\hat{\kappa}_2$	−2.26011	0.17580
$\hat{\alpha}_3$	−0.06323	0.02830	$\hat{\kappa}_3$	−1.54132	0.15572
$\hat{\beta}_1$	0.00215	0.01015	$\hat{\gamma}_1$	−0.01914	0.03492
$\hat{\beta}_2$	−0.00915	0.00482	$\hat{\gamma}_2$	0.04075	0.01798
$\hat{\beta}_3$	0.09460	0.03338	$\hat{\gamma}_3$	−0.34234	0.14195
$\hat{\beta}_4$	−0.06811	0.03850	$\hat{\gamma}_4$	0.45315	0.17028
$\hat{\beta}_5$	−0.07810	0.05148	$\hat{\gamma}_5$	0.72040	0.20844
$\hat{\beta}_6$	0.070018	0.06933	$\hat{\gamma}_6$	−0.63781	0.27353
$\ell(MLE)$	−467.25			−463.7662	

TABLE 5 Maximum likelihood estimate (MLE) by ppm for Example 4.4

MLE	ppm	SE	MLE	ppm	SE
$\hat{\alpha}_1$	−0.33172	0.03760	$\hat{\kappa}_1$	−2.73868	0.19815
$\hat{\alpha}_2$	−0.12300	0.02957	$\hat{\kappa}_2$	−2.27915	0.17772
$\hat{\alpha}_3$	−0.05992	0.02706	$\hat{\kappa}_3$	−1.55364	0.15747
$\hat{\beta}_1$	0.00028	0.00987	$\hat{\gamma}_1$	−0.01757	0.03504
$\hat{\beta}_2$	−0.00913	0.00461	$\hat{\gamma}_2$	0.04061	0.01799
$\hat{\beta}_3$	0.10509	0.03370	$\hat{\gamma}_3$	−0.35545	0.14311
$\hat{\beta}_4$	−0.08044	0.03841	$\hat{\gamma}_4$	0.47034	0.17204
$\hat{\beta}_5$	−0.08981	0.05089	$\hat{\gamma}_5$	0.73605	0.20977
$\hat{\beta}_6$	0.08464	0.06810	$\hat{\gamma}_6$	−0.65454	0.27431
ℓ(MLE)	−464.1204			−460.1224	

TABLE 6 Maximum likelihood estimate (MLE) by ppm for Example 4.4

and in this example, significance. Clearly, in practice, the outcome order should be determined prior to beginning any analysis.

We now briefly explore the implications of certain data removals. We removed the individuals (230 and 348) with the following covariates: $x_1 = -1.73759$, $x_2 = 3.01923$, $x_3 = 1$, $x_4 = 0$, $x_5 = 0$. The analysis conducted before is repeated and the MLEs are in Table6. Of note is that the MLEs in Tables 5 and 6 are very similar. However, once individuals 230 and 348 are removed the MLEs for both the reverse and not reverse coded are now both in the interior. By removing the data points associated with the boundary MLE, we find that the MLE is now in the interior of the parameter space. While such data deletion would not be done in practice, we use it here to demonstrate that the MLE can move from the boundary to the interior of the parameter space. In the next Section 5, we continue the exploration of characteristics that may determine the MLE to be on a boundary.

4.5 | Application 4: Mental impairment dataset

We now consider an example where the MLE is considered on a boundary. Consider the Mental Impairment data presented in table 7.5 of Agresti.[39] This involves 40 subjects with outcome ordinal variable Mental Impairment (1 - impaired, 2 - moderate, 3 - mild, 4 - well) and binary covariates : Socioeconomic status (x_2: high=1,low=0) and Life Event (x_1) a composite score over the last 3 years. Using this data, we will explore the following model: $P[y \leq j|x_1, x_2] = \alpha_j + \beta_1 x_1 + \beta_2 x_2$. `ppm` estimates the maximum likelihood to occur at $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_1, \hat{\beta}_2)' = (-1.68422, -1.10888, -0.54929, 0.09155, -0.42750)'$. The determination of an MLE on a boundary is using the rule that if at least one fitted cumulative probability is greater than 0.9999. `ppm` determines that for $j = 3$ individuals with $x_1 = 6$ and

TABLE 7 Fraction of boundary maximum likelihood estimate (MLE) for certain $n_{kj} = 0$ for Example 5.1

Only one $n_{kj} = 0$	**Boundary MLE %**
$n_{11} = 0$	0 %
$n_{12} = 0$	0 %
$n_{13} = 0$	100 %
$n_{21} = 0$	0%
$n_{22} = 0$	0 %
$n_{23} = 0$	100%
Two $n_{kj} = 0$	**Boundary MLE %**
$n_{12} = 0$	0 %
$n_{13} = 0$	
$n_{22} = 0$	100 %
$n_{23} = 0$	

$x_2 = 0$ have a fitted cumulative probability ($\hat{P}[y \le 3|x_1 = 6, x_2 = 0]$) close to 1. Since the MLE is then assumed to be on a boundary, the estimates of the SEs are not directly available.

We also note that if the outcome was reversely coded: 4 - impaired, 3 - moderate, 2 - mild, 1 - well). The model could be rewritten as $\log[1 - P(y \le j|x_1, x_2)] = \kappa_j + \gamma_1 x_1 + \gamma_2 x_2$. ppm estimates the maximum likelihood to occur at $(\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, \hat{\gamma}_1, \hat{\gamma}_2)' = (-0.94908, -0.25593, -2.1611e-12, -0.07441, 0.07620)'$. The determination of an MLE on a boundary is using the rule that if at least one fitted cumulative probability is greater than 0.9999. ppm determines that for $j = 3$ individuals with $x_1 = 0$ and $x_2 = 0$ have a fitted cumulative probability ($\hat{P}[y \le 3|x_1 = 0, x_2 = 0]$) close to 1. Since the MLE is then assumed to be on a boundary, the estimates of the SEs are not directly available. Interestingly, the MLE is still on a boundary regardless of the outcome coding order. In the next Section 5, MLE on a boundary will be explored further.

5 | MLE ON THE BOUNDARY

This section provides a conjecture about the location of an MLE on a boundary for a PPM. A boundary for a PPM is $\{(\boldsymbol{\alpha},\boldsymbol{\beta})|\alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta} = \mathbf{0}\} \cup \{\boldsymbol{\alpha}|\alpha_{j+1} = \alpha_j\}$.

The following examples presented are based on the second expression for CFE in Theorem 2 when $n_{kJ} = 0$ for some k. The first Example 5.1, demonstrates that a boundary MLE can occur when $n_{kJ} = 0$. Then a second Example 5.2 demonstrates the CFE expressions. In simulating several conditions based on values of n_{kj}, it is noted that if $n_{kJ} = 0$ for a particular group k then the MLE is sometimes on a boundary. This has the intuition that there is no additional information for group k having category value J, that is $P(y_k \le (J-1)) = 1$. The exact conditions for a boundary MLE are not known to us at this time, but the examples presented in this section focus on settings when $n_{kJ} = 0$.

5.1 | Example 1: Simulation with $n_{kj} = 0$

Recall Example 4.1, a model with $\log[P(y \le j|x)] = \alpha_j + \beta_1 x$ where $x = \{0, 1\}$ having two cuts. Datasets were constructed by taking n_{kl} from vector $(0, 1, 2, 5, 10, 100)'$ for $k = \{1, 2\}$ and $l = \{1, 2, 3\}$. For each dataset, the MLE were computed by ppm (using default controls in ppm) and the determination of an MLE on a boundary was done by noting if the fitted probabilities were greater than 0.9999. The percentage of scenarios with boundary MLE was determined for scenarios with only one $n_{kl} = 0$ and two $n_{kl} = 0$. The results are presented in Table 7. Initially, it would appear that when $n_{kJ} = 0$ for only one k, there is always a boundary MLE (see $n_{13} = 0$ and $n_{23} = 0$). However, when both $n_{12} = 0$ and $n_{13} = 0$, there are no boundary MLE, but for both $n_{22} = 0$ and $n_{23} = 0$ there is always a boundary MLE.

5.2 | Example 2: Simulation with $n_{kJ} = 0$

Suppose a model where $\log[P(y \le j|x, z)] = \alpha_j + \beta_1 x + \beta_2 z$ where $x = \{0, 1\}$, $z = \{0, 1\}$ and $j = \{1, 2\}$ cuts. Datasets were constructed by taking n_{kl} and n_{43} from vector $(10, 100)'$ for $k = \{1, 2, 3, 4\}$ and $l = \{1, 2\}$ and drawing n_{k3} from vector

TABLE 8 Fraction of Boundary maximum likelihood estimate (MLE) for certain $n_{k3} = 0$ for Example 5.2

which $n_{k3} = 0$	Number on boundary	Number of scenarios	Percent on boundary (%)
only $n_{13} = 0$	362	512	70.70
only $n_{23} = 0$	415	512	81.05
only $n_{33} = 0$	415	512	81.05
only $n_{13} = n_{23} = 0$	512	512	100.00
only $n_{13} = n_{33} = 0$	512	512	100.00
only $n_{23} = n_{33} = 0$	320	512	62.50
$n_{13} = n_{23} = n_{33} = 0$	512	512	100.00
$n_{13}, n_{23}, n_{33} \neq 0$	0	512	0.00

$\{0, 100\}$ for $k = \{1, 2, 3\}$. For each dataset, the MLE were computed by `ppm` (using default controls in `ppm`) and the determination of an MLE on a boundary was done by noting if the fitted probabilities were greater than 0.9999.

$$\boldsymbol{C}|\boldsymbol{N} = \begin{bmatrix} 0 & 0 & n_{11} & n_{12} & n_{13} \\ 1 & 0 & n_{21} & n_{22} & n_{23} \\ 0 & 1 & n_{31} & n_{32} & n_{33} \\ 1 & 1 & n_{41} & n_{42} & n_{43} \end{bmatrix}.$$

The percentage of scenarios with boundary MLE was determined for scenarios with only one $n_{k3} = 0$, only two $n_{k3} = 0$ and only three $n_{k3} = 0$ are found in Table 8. Unlike Example 5.1, the datasets here have the fraction of the MLE on the boundary between 0% and 100%. This implies that it is not just $n_{kJ} = 0$ that induces an MLE to be on a boundary. The simulations suggest that if $n_{kJ} \neq 0$ then an MLE will not be on a boundary. However, additional investigation is required.

As a final note from this example, CFE from 2 in Theorem 2 can be compared with `ppm` for MLE in the interior. In the setting when only $n_{13} = 0$, there are 150 scenarios of values for n_{kj} for which the MLE is in the interior. All 150 datasets were constructed and the MLE using CFE and `ppm` were calculated. The absolute difference in `ppm` and CFE MLE for each parameter was on average (9.060e-07, 7.889e-07, 5.984e-07, 6.611e-07) with minimum absolute difference (6.626e-08, 1.069e-08, 6.561e-08, 1.565e-08) and maximum absolute difference (2.844e-06, 3.243e-06, 2.206e-06, 2.607e-06). Several results are presented in Table 9. Note that CFE used:

$$\boldsymbol{C}^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{C}^*_{-\boldsymbol{I}} = \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The MLEs are given by

$$\hat{\alpha}_1 = \hat{\alpha}_2 - \log\left(\frac{n_{11} + n_{21} + n_{31} + n_{41} + n_{12} + n_{22} + n_{32} + n_{42}}{n_{11} + n_{21} + n_{31} + n_{41}}\right) \tag{8}$$

$$\hat{\alpha}_2 = \log\left(\frac{n_{11} + n_{12} + n_{21} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22} + n_{23}}\right) + \log\left(\frac{n_{11} + n_{12} + n_{31} + n_{32}}{n_{11} + n_{12} + n_{31} + n_{32} + n_{33}}\right) - \log\left(\frac{n_{41} + n_{42} - n_{11} - n_{12}}{n_{43} + n_{41} + n_{42} - n_{11} - n_{12}}\right) \tag{9}$$

$$\hat{\beta}_1 = \log\left(\frac{n_{41} + n_{42} - n_{11} - n_{12}}{n_{43} + n_{41} + n_{42} - n_{11} - n_{12}}\right) - \log\left(\frac{n_{11} + n_{12} + n_{31} + n_{32}}{n_{11} + n_{12} + n_{31} + n_{32} + n_{33}}\right) \tag{10}$$

$$\hat{\beta}_2 = \log\left(\frac{n_{41} + n_{42} - n_{11} - n_{12}}{n_{43} + n_{41} + n_{42} - n_{11} - n_{12}}\right) - \log\left(\frac{n_{11} + n_{12} + n_{21} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22} + n_{23}}\right) \tag{11}$$

TABLE 9 Maximum likelihood estimate (MLE) by closed form expression (CFE) and ppm for Example 5.2

(n_{11}, n_{12}, n_{13}) (n_{21}, n_{22}, n_{23}) (n_{31}, n_{32}, n_{33}) (n_{41}, n_{42}, n_{43})	MLE	CFE	ppm
(100, 10, 0)	$\hat{\alpha}_1$	−1.02761	−1.02761
(100, 100, 100)	$\hat{\alpha}_2$	−0.45381	−0.45381
(100, 100, 100)	$\hat{\beta}_1$	0.17422	0.17423
(100, 100, 10)	$\hat{\beta}_2$	0.17422	0.17423
(10, 10, 0)	$\hat{\alpha}_1$	−3.58975	−3.58975
(10, 10, 100)	$\hat{\alpha}_2$	−1.71795	−1.71795
(10, 100, 100)	$\hat{\beta}_1$	0.46518	0.46519
(10, 100, 10)	$\hat{\beta}_2$	1.14740	1.14740
(10, 10, 0)	$\hat{\alpha}_1$	−1.87877	−1.87877
(10, 10, 100)	$\hat{\alpha}_2$	−1.18562	−1.18562
(100, 100, 100)	$\hat{\beta}_1$	−0.06714	−0.06714
(100, 100, 100)	$\hat{\beta}_2$	0.81093	0.81093
(10, 100, 0)	$\hat{\alpha}_1$	−2.02613	−2.02613
(10, 10, 100)	$\hat{\alpha}_2$	−1.03573	−1.03573
(10, 10, 100)	$\hat{\beta}_1$	0.46518	0.46518
(100, 100, 10)	$\hat{\beta}_2$	0.46518	0.46518

6 | CONCLUDING REMARKS

This manuscript expands the understanding and methodology of the PPM for use by a data analyst. The condition for the strict concavity of the log-likelihood requires assessment of the rank of $\boldsymbol{C}^*_{-\boldsymbol{I}}$. An analyst should consider the potential of multicollinearity and an approach to its resolution.

While the proportional odds model is commonly used, the resulting estimates have interpretations in terms of functions of the log odds such as odds ratios. There has been significant discussion in the literature on misinterpretations of odds as probabilities. The PPM gives estimates of the log probability (like rate ratios, risk ratios, and health ratios). These estimates are typically of interest to analysts in the health sciences. The trade off for the PPM providing such estimates is that the log link implicitly introduces linear inequality constraints on the parameter space. We feel that since the function `ppm` uses a constrained optimization methodology of `constrOptim`, the MLE can be determined reliably in the interior and on the boundary of the parameter space. This was shown through several examples. Having CFEs for the MLE also permitted the comparison with the MLE determined by `ppm`.

As is noted, additional investigation is required for the exact conditions for an MLE on a boundary. However, having groups for which $n_{kJ} = 0$ appears to play a role for an MLE on a boundary. The use of measured explanatory variables may introduce some groups with $n_{kJ} = 0$. Another consideration is the use of the covariance matrix from the inverse of the negative observed Hessian when an MLE is on a boundary. The analyst should assess the behavior of the Hessian near a boundary. In any case, the interpretation of such a covariance matrix may be limited and its utility might be misleading. We present the assessment of the proportionality assumption using the score test. Additional exploration of likelihood and Wald-type tests of proportionality is required.

For ordinal outcomes, the choice to use the log link is now available, and this choice can be based on the study design, the measure of association and the interpretation(s) desired. We have shown the utility of the function `ppm` in our preliminary version of the R package `lcpm`.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

ORCID

Gurbakhshash Singh https://orcid.org/0000-0003-4760-6395

REFERENCES

1. Aitchison J, Silvey SD. The generalization of probit analysis to the case of multiple responses. *Biometrika*. 1957;44(1):131-140.
2. McCullagh P. Regression models for ordinal data. *J Royal Stat Soc Ser B (Methodol)*. 1980;42(2):109-142.
3. Blizzard CL, Quinn SJ, Canary JD, Hosmer DW. Log-link regression models for ordinal responses. *Open J Stat*. 2013;3:16-25.
4. Williams R. Fitting heterogeneous choice models with OGLM. *Stata J*. 2010;10(4):540-567. http://www.stata-journal.com/article.html?article=st0208.
5. Williams R. OGLM help file. http://fmwww.bc.edu/RePEc/bocode/o/oglm.html. Accessed April 23, 2018.
6. R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing; 2018. Vienna, Austria. https://www.R-project.org/.
7. Yee TW. VGAM: vector generalized linear and additive models, R package version 1.0-4. 2017. https://CRAN.R-project.org/package=VGAM. Accessed April 23, 2018.
8. Singh G, Fick GH. LCPM: ordinal outcomes: generalized linear models with the log link. R package version 0.1.0. 2019. https://CRAN.R-project.org/package=lcpm. Accessed January 28, 2019.
9. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol*. 1986;123(1):174-184.
10. Schouten EG, Dekker J, Kok FJ, et al. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Stat Med*. 1993;12(18):1733-1745.
11. Zhang J, Yu K. What's the relative risk? a method of correcting the odds ratio in cohort studies of common outcomes. *J Am Med Assoc*. 1998;280(19):1690-1691.
12. Skov T, Deddens J, Petersen MR, Endahl L. Prevalence proportion ratios: estimation and hypothesis testing. *Int J Epidemiol*. 1998;27(1):91-95.
13. Deddens J, Petersen MR, Lei X. Estimation of prevalence ratios when PROC GENMOD does not converge. Paper presented at: Proceedings of the 28th Annual SAS Users Group International Conference; 2003.
14. Deddens J, Petersen MR. Maximum likelihood estimation of the log-binomial model. *Commun Stat Theory Methods*. 2010;39:874-883.
15. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;154:702-706.
16. Carter R, Lipsitz S, Tilley B. Quasi-likelihood estimation for relative risk regression models. *Biostatistics*. 2005;6(1):39-44.
17. Lumley T, Kronmal R, Ma S. *Relative Risk Regression in Medical Research: Models, Contrasts, Estimators, and Algorithms*. UW Biostatistics Working Paper Series. Washington, DC: The Berkeley Electronic Press; 2006.
18. Yu B, Wang Z. Estimating relative risks for common outcome using PROC NLP. *Comput Methods Prog Biomed*. 2008;90:179-186.
19. Austin P. Absolute risk reductions, relative risks, relative risk reductions, and number needed to treat can be obtained from a logistic regression model. *J Clin Epidemiol*. 2010;63(1):2-6.
20. Chu H, Cole SR. Estimation of risk ratios in cohort studies with common outcomes: a bayesian approach. *Epidemiology*. 2010;21(6):855-862.
21. Savu A, Liu Q, Yasui Y. Estimation of relative risk and prevalence ratio. *Stat Med*. 2010;29(22):2269-2281.
22. Marschner IC, Gillett AC. Relative risk regression: reliable and flexible methods for log-binomial models. *Biostatistics*. 2012;13(1):179-192.
23. Dwivedi AK, Mallawaarachchi I, Lee S, Tarwater P. Methods for estimating relative risk in studies of common binary outcomes. *J Appl Stat*. 2014;41(3):484-500.
24. Zhou R, Sivaganesan S, Longla M. An objective Bayesian estimation of parameters in a log-binomial model. *J Stat Planning Infer*. 2014;146:113-121.
25. Luo J, Zhang J, Sun H. Estimation of relative risk using a log-binomial model with constraints. *Comput Stat*. 2014;29(5):981-1003.
26. Salmeron D, Cano JA, Chirlaque MD. Reducing Monte Carlo error in the Bayesian estimation of risk ratios using log-binomial regression models. *Stat Med*. 2015;34:2755-2767.
27. Fitzmaurice GM, Lipsitz SR, Arriaga A, Sinha D, Greenberg C, Gawande AA. Almost efficient estimation of relative risk regression. *Biostatistics*. 2014;15:745-756.
28. Marschner IC. Relative risk regression for binary outcomes: methods and recommendations. *Aust NZ J Stat*. 2015;57(4):437-462.
29. de Andrade BB, de Leon Andrade JM. Some results for maximum likelihood estimation of adjusted relative risks. *Commun Stat Theory Methods*. 2017;47(4):1-20. https://doi.org/10.1080/03610926.2017.1402045.
30. Lange K. An adaptive barrier method for convex programming. *Methods Appl Anal*. 1994;1:392-402.
31. Lange K. *Optimization*. 2nd ed. New York, NY: Springer; 2004.
32. Liu Q, Shepherd BE, Li C, Harrell FE Jr. Modeling continuous response variables using ordinal regression. *Stat Med*. 2017;36:4316-4335.
33. Efron B, Hinkley D. Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*. 1978;65:457-487.
34. Wedderburn RWM. On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*. 1976;63(1):27-32.
35. Rockafellar RT. *Convex Analysis*. Princeton, NJ: Princeton University Press; 1970.
36. Hill HA, Coates RJ, Austin H, et al. Racial differences in tumor grade among women with endometrial cancer. *Gynecol Oncol*. 1995;56:154-163.

37. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.
38. Braglia L. Aplore3: datasets from Hosmer Lemeshow and Sturdivant. *Applied Logistic Regression* 3rd ed; 2013. R package version 0.9. 2016. https://CRAN.R-project.org/package=aplore3. Accessed October 26, 2019.
39. Agresti A. *Categorical Data Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2002.
40. Berkovitz L. *Convexity and Optimization in Rn*. New York, NY: John Wiley & Sons Inc; 2002.
41. Peterson B, Harrell FE. Partial proportional odds models for ordinal response variables. *Appl Stat*. 1990;39:313-329.
42. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*. 1990;46:1171-1178.
43. SAS/STAT(R) 9.3 User's Guide. https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_logistic_sect039.htm. Accessed August 16, 2017.

How to cite this article: Singh G, Hilton Fick G. Ordinal outcomes: A cumulative probability model with the log link and an assumption of proportionality. *Statistics in Medicine*. 2020;1–19. https://doi.org/10.1002/sim.8479

APPENDIX A

A.1 PPM proof of uniqueness of the MLE

The proof of the conditions for strict concavity for the log-likelihood function for the PPM is as follows:

Proof of Theorem 1. Begin by noting that Ω_{PPM} is a convex set. This is easily established noting for the PPM, the parameter space is $\Omega_{\text{PPM}} = \{(\boldsymbol{\alpha}, \boldsymbol{\beta})|\alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta} \leq \boldsymbol{0}, \alpha_{j+1} \geq \alpha_j\}$. Suppose $\boldsymbol{\epsilon_1} \in \Omega_{PPM}$, $\boldsymbol{\epsilon_2} \in \Omega_{\text{PPM}}$ and constant $\gamma \in [0, 1]$. Let $\boldsymbol{\epsilon}^* = \gamma\boldsymbol{\epsilon_1} + (1-\gamma)\boldsymbol{\epsilon_2}$ and

$$\begin{aligned}\boldsymbol{X\epsilon}^* &= \boldsymbol{X}(\gamma\boldsymbol{\epsilon_1} + (1-\gamma)\boldsymbol{\epsilon_2})\\ &= \gamma\underbrace{\boldsymbol{X\epsilon_1}}_{\leq\boldsymbol{0}} + (1-\gamma)\underbrace{\boldsymbol{X\epsilon_2}}_{\leq\boldsymbol{0}}\\ &\leq \boldsymbol{0},\end{aligned}$$

$\boldsymbol{\epsilon}^* \in \Omega_{\text{PPM}}$ and Ω_{PPM} is a convex set.

To establish the strictly concave in the interior of the parameter space, it suffices to show that the Hessian is negative definite $\boldsymbol{z'Hz} < 0$ for all $\boldsymbol{z} \neq \boldsymbol{0}$. Define $\boldsymbol{z'} = (\boldsymbol{\xi'}, \boldsymbol{\zeta'}) = (\underbrace{\xi_1, \xi_2, \dots, \xi_{J-1}}_{\xi'}, \underbrace{\zeta_1, \zeta_2, \dots, \zeta_p}_{\zeta'})$ and

$$\begin{aligned}\boldsymbol{z'Hz} &= [\boldsymbol{\xi'\zeta'}]\begin{bmatrix}\boldsymbol{H}_{\alpha\alpha} & \boldsymbol{H}_{\alpha\beta}\\ \boldsymbol{H}'_{\alpha\beta} & \boldsymbol{H}_{\beta\beta}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix}\\ &= \underbrace{\boldsymbol{\xi'H}_{\alpha\alpha}\boldsymbol{\xi}}_{(1)} + \underbrace{\boldsymbol{\xi'H}_{\alpha\beta}\boldsymbol{\zeta}}_{(2)} + \underbrace{\boldsymbol{\zeta'H}'_{\alpha\beta}\boldsymbol{\xi}}_{(3)} + \underbrace{\boldsymbol{\zeta'H}_{\beta\beta}\boldsymbol{\zeta}}_{(4)}.\end{aligned} \tag{A1}$$

Now assess each term and recall that $\delta_k = n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}$ and $\gamma_{k,j+1} = n_{k(j+1)}\frac{\exp(\alpha_{j+1}+\alpha_j)}{(\exp(\alpha_{j+1})-\exp(\alpha_j))^2}$:

(1) Assessment of the first term in (A1):

$$\begin{aligned}\boldsymbol{\xi'H}_{\alpha\alpha}\boldsymbol{\xi} &= \sum_{l=1}^{J-1} A_{ll}\xi_l^2 + 2\sum_{q=1}^{J-2} A_{(q+1)q}\xi_{q+1}\xi_q\\ &= -\sum_{l=2}^{J-2}\sum_{k=1}^{m}(\gamma_{k,l+1} + \gamma_{k,l})\xi_l^2 - \sum_{k=1}^{m}(\delta_k + \gamma_{k,J-1})\xi_{J-1}^2 - \sum_{k=1}^{m}\gamma_{k,2}\xi_1^2 + 2\sum_{q=1}^{J-2}\sum_{k=1}^{m}\gamma_{k(q+1)}\xi_{q+1}\xi_q\\ &= -\sum_{k=1}^{m}\delta_k\xi_{J-1}^2 - \sum_{k=1}^{m}\left[\sum_{l=2}^{J-2}(\gamma_{k,l+1} + \gamma_{k,l})\xi_l^2 + \gamma_{k,J-1}\xi_{J-1}^2 + \gamma_{k,2}\xi_1^2 - 2\sum_{q=1}^{J-2}\gamma_{k(q+1)}\xi_{q+1}\xi_q\right]\end{aligned}$$

$$= -\sum_{k=1}^{m}\delta_k\xi_{J-1}^2 - \sum_{k=1}^{m}\left[\sum_{r=1}^{J-2}\gamma_{k,r+1}(\xi_r-\xi_{r+1})^2\right]$$

$$= -\sum_{k=1}^{m} n_{kJ}\xi_{J-1}^2 \underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0} - \sum_{k=1}^{m}\sum_{r=1}^{J-2} n_{k(r+1)}\underbrace{\frac{\exp(\alpha_{r+1}+\alpha_r)}{(\exp(\alpha_{r+1})-\exp(\alpha_r))^2}}_{>0}(\xi_r-\xi_{r+1})^2. \tag{A2}$$

(2) Assessment of the second and third terms in (A1):

$$\begin{aligned}\boldsymbol{\xi}'\boldsymbol{H}_{\alpha\beta}\boldsymbol{\zeta} &= \boldsymbol{\zeta}'\boldsymbol{H}_{\alpha\beta}'\boldsymbol{\xi}\\ &= -\xi_{J-1}\sum_{l=1}^{p}\sum_{k=1}^{m} n_{kJ}c_{kl}\zeta_l \underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0}.\end{aligned} \tag{A3}$$

(3) For the last term in (A1), denote $\delta_k = n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}$ gives:

$$\begin{aligned}\boldsymbol{\zeta}'\boldsymbol{H}_{\beta\beta}\boldsymbol{\zeta} &= \sum_{l=1}^{p} B_{ll}\zeta_l^2 + 2\sum_{l<q}\sum_{q=1}^{p} B_{lq}\zeta_l\zeta_q\\ &= -\sum_{k=1}^{m}\delta_k\left[\sum_{l=1}^{p} c_{kl}^2\zeta_l^2 + 2\sum_{l<q}\sum_{q=1}^{p} c_{kl}c_{kq}\zeta_l\zeta_q\right]\\ &= -\sum_{k=1}^{m}\delta_k\left[\sum_{l=1}^{p} c_{kl}\cdot\zeta_l\right]^2\\ &= -\sum_{k=1}^{m} n_{kJ}\underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0}\left(\sum_{l=1}^{p} c_{kl}\cdot\zeta_l\right)^2.\end{aligned} \tag{A4}$$

Now recombine the terms (A2), (A3), and (A4) and rearrange them as

$$\begin{aligned}\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} = &-\sum_{k=1}^{m}\left[\sum_{r=1}^{J-2} n_{k(r+1)}\frac{\exp(\alpha_{r+1}+\alpha_r)}{(\exp(\alpha_{r+1})-\exp(\alpha_r))^2}(\xi_r-\xi_{r+1})^2\right]\\ &-\sum_{k=1}^{m} n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}\left[\xi_{J-1}^2 + 2\xi_{J-1}\sum_{l=1}^{p} c_{kl}\zeta_l + \left(\sum_{l=1}^{p} c_{kl}\cdot\zeta_l\right)^2\right].\end{aligned}$$

Factoring this gives:

$$\begin{aligned}\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} = &-\sum_{k=1}^{m}\left[\sum_{r=1}^{J-2} n_{k(r+1)}\frac{\exp(\alpha_{r+1}+\alpha_r)}{(\exp(\alpha_{r+1})-\exp(\alpha_r))^2}(\xi_r-\xi_{r+1})^2\right]\\ &-\sum_{k=1}^{m} n_{kJ}\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}\left(\xi_{J-1} + \sum_{l=1}^{p} c_{kl}\zeta_l\right)^2.\end{aligned} \tag{A5}$$

From (A5), $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} \le 0$ for all $\boldsymbol{z}$ and $\boldsymbol{H}$ is negative semidefinite and the log-likelihood is concave.[40] However, additional conditions can be provided for which the log-likelihood is strictly concave. Recall that $\boldsymbol{N}^{(-J)}$ has $n^{(j)} > 0$ for $j = 1, \dots, J-1$. Let $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} = T_1 + T_2$ where $T_1 \le 0$, $T_2 \le 0$,

$$T_1 = -\sum_{k=1}^{m}\left[\sum_{r=1}^{J-2} \underbrace{n_{k(r+1)}}_{\geq 0} \underbrace{\frac{\exp(\alpha_{r+1}+\alpha_r)}{(\exp(\alpha_{r+1})-\exp(\alpha_r))^2}}_{>0} \underbrace{(\xi_r - \xi_{r+1})^2}_{\geq 0}\right],$$

and

$$T_2 = -\sum_{k=1}^{m} \underbrace{n_{kJ}}_{\geq 0} \underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0} \underbrace{\left(\xi_{J-1}+\sum_{l=1}^{p} c_{kl}\zeta_l\right)^2}_{\geq 0}.$$

1. Suppose $n_{kJ} = 0$ for $k \in I$ where $m - |I| \geq (p+1)$ and suppose $\boldsymbol{C}^*_{-\boldsymbol{I}}$ is full column rank. For $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} = T_1 + T_2$ to sum to 0, both T_1 and T_2 must be zero.

$$T_2 = -\sum_{k\notin I} \underbrace{n_{kJ}}_{>0} \underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0} \left(\xi_{J-1}+\sum_{l=1}^{p} c_{kl}\zeta_l\right)^2$$

For T_2 to be zero for $k \notin I$, $\xi_{J-1} + \sum_{l=1}^{p} c_{kl}\zeta_l = 0$. This represents rows of the system $\boldsymbol{C}^*_{-\boldsymbol{I}}\boldsymbol{\epsilon} = \boldsymbol{0}$, where $\boldsymbol{\epsilon}' = (\xi_{J-1}, \boldsymbol{\zeta}')$. $\boldsymbol{C}^*_{-\boldsymbol{I}}$ is full column rank and the solution to the system is $\boldsymbol{\epsilon} = \boldsymbol{0}$.

$$T_1 = -\sum_{k=1}^{m}\left[\sum_{r=1}^{J-2} \underbrace{n_{k(r+1)}}_{\geq 0} \underbrace{\frac{\exp(\alpha_{r+1}+\alpha_r)}{(\exp(\alpha_{r+1})-\exp(\alpha_r))^2}}_{>0} \underbrace{(\xi_r - \xi_{r+1})^2}_{\geq 0}\right]$$

For T_1 to be zero, $n_{k(r+1)} = 0$ or $(\xi_r - \xi_{r+1})^2 = 0$ or both. From the construction of $\boldsymbol{N}$, the row total $(n_{(k)} > 0)$ and column totals $(n^{(j)} > 0)$ are nonzero. Thus, for each $j = 1, \ldots, J-2$, there is a k for which $n_{k(j+1)} > 0$. For these k, $\xi_r - \xi_{r+1} = 0$ for $r = 1, \ldots, J-2$ and $\xi_r = \xi_{r+1}$. From $T_2 = 0$, $\xi_{J-1} = 0$ and working backward recursively, $\xi_r = 0$ for $r = 1, \ldots, J-1$. Finally, $\boldsymbol{\xi} = \boldsymbol{0}$ and $T_1 = 0$.

$\boldsymbol{\xi}'\boldsymbol{H}_{\alpha\alpha}\boldsymbol{\xi} + \boldsymbol{\xi}'\boldsymbol{H}_{\alpha\beta}\boldsymbol{\zeta} + \boldsymbol{\zeta}'\boldsymbol{H}'_{\alpha\beta}\boldsymbol{\xi} + \boldsymbol{\zeta}'\boldsymbol{H}_{\beta\beta}\boldsymbol{\zeta} < 0$ with equality when $\boldsymbol{\xi} = \boldsymbol{0}$ and $\boldsymbol{\zeta} = \boldsymbol{0}$. Combining all the information, it is clear that $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} < 0$ with equality when $\boldsymbol{z} = \boldsymbol{0}$. $\boldsymbol{H}$ is negative definite and the log-likelihood function is strictly concave.[40]

2. Suppose $n_{kJ} = 0$ for $k \in I$ and $m - |I| < (p+1)$. For $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} = T_1 + T_2$ to sum to 0, both T_1 and T_2 must be zero.

$$T_2 = -\sum_{k\notin I} \underbrace{n_{kJ}}_{>0} \underbrace{\frac{\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta})}{(1-\exp(\alpha_{J-1}+\boldsymbol{c}_k'\boldsymbol{\beta}))^2}}_{>0} \left(\xi_{J-1}+\sum_{l=1}^{p} c_{kl}\zeta_l\right)^2$$

For T_2 to be zero for $k \notin I$, $\xi_{J-1} + \sum_{l=1}^{p} c_{kl}\zeta_l = 0$. However, the number of rows of the system $\boldsymbol{C}^*_{-\boldsymbol{I}}\boldsymbol{\epsilon} = \boldsymbol{0}$ are fewer than the number of columns. There is a nontrivial solution to the system and $\boldsymbol{\epsilon} \neq \boldsymbol{0}$. Combining all the information, it is clear that $\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z} \leq 0$ for all $\boldsymbol{z}$. $\boldsymbol{H}$ is negative semi-definite and the log-likelihood function is concave.[40]

■

A.2 Appendix vglm

`vglm` is a function that is part of the `VGAM` package in R. This function attempts to determine the MLE with a cumulative probability model for ordinal outcomes with or without a proportionality assumption and with several different links. This function uses an unconstrained iteratively reweighted least squares algorithm to try to determine the MLE. This algorithm may yield incorrect supposed MLE outside of Ω_{PPM}. `vglm` function suboption

TABLE A1 Maximum likelihood estimate (MLE) by closed form expression (CFE) and ppm and the apparent maximum by vglm for Example 4.1

(n_{11}, n_{12}, n_{13}) (n_{21}, n_{22}, n_{23})	MLE	CFE	ppm	vglm
(1, 1, 1)	$\hat{\alpha}_1$	−1.32176	−1.32175	−1.32176
(1, 2, 10)	$\hat{\alpha}_2$	−0.40547	−0.40546	−0.40547
	$\hat{\beta}_1$	−1.06087	−1.06087	−1.06087
(1, 1, 5)	$\hat{\alpha}_1$	−1.94591	−1.94591	0.674796[a]
(2, 2, 100)	$\hat{\alpha}_2$	−1.25276	−1.25276	1.36794[a]
	$\hat{\beta}_1$	−2.00533	−2.00533	−4.57417[a]
(1, 1, 2)	$\hat{\alpha}_1$	−1.73460	−1.73460	−0.48736[a]
(5, 10, 100)	$\hat{\alpha}_2$	−0.69315	−0.69314	0.55409[a]
	$\hat{\beta}_1$	−1.34373	−1.34374	−2.58696[a]
(1, 1, 1)	$\hat{\alpha}_1$	−1.09861	−1.09861	0.31298[a]
(10, 10, 100)	$\hat{\alpha}_2$	−0.40547	−0.40547	1.00613[a]
	$\hat{\beta}_1$	−1.38629	−1.38629	−2.79545[a]
(1, 1, 5)	$\hat{\alpha}_1$	−1.94591	−1.94591	−1.94591
(100, 100, 10)	$\hat{\alpha}_2$	−1.25276	−1.25276	−1.25276
	$\hat{\beta}_1$	1.20397	1.20397	1.20397
(100, 1, 5)	$\hat{\alpha}_1$	−0.07523	−0.07523	0.15419[a]
(10, 2, 100)	$\hat{\alpha}_2$	−0.04832	−0.04832	0.18110[a]
	$\hat{\beta}_1$	−2.18527	−2.18527	−1.63150[a]
(100, 100, 100)	$\hat{\alpha}_1$	−1.09861	−1.09861	−1.09861
(100, 100, 100)	$\hat{\alpha}_2$	−0.40547	−0.40547	−0.40547
	$\hat{\beta}_1$	0	6.19e-08	−3.4613e-16

[a]MLE from vglm is outside parameter space.

`cumulative(parallel=TRUE,link="loge")` can be used to try to estimate the MLE for a PPM but we cannot recommend its use.

Recall Example 4.1. For many scenarios, the absolute difference in `vglm` and CFE MLE were comparable to the results comparing `ppm` and CFE MLE. However, for `vglm`, 35 of the first 381 datasets had large differences. Often with an incorrect maximum [that is not the MLE] being outside the constrained parameter space. Some examples of this are presented in scenarios in Table A1.

This example compared the closed form MLE with the MLE from `ppm`. The differences were very small. However, `vglm` converged outside the constrained parameter space even though the MLE is in the interior of the constrained parameter space. At present, the conditions have not been determined for which `vglm` finds a maximum (that is not the MLE) outside of the constrained parameter space but `ppm` finds the actual MLE within the constrained parameter space.

A.3 Appendix Score Test of Proportionality

In ordinal logistic regression, the assessment of the proportional odds assumption can be performed with the likelihood ratio test,[41] the Wald (or Brant) test,[42] or the score test.[43] In this section, we use a score test that is analogous to the test used with ordinal logistic regression.[43] Additional information for this test can be found in Blizzard et al.[3]

The PPM from Section 2 makes the assumption that different levels of the ordinal outcomes share a common slope β. This can be tested using the hypothesis: $H_o\ :\ \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \ldots = \boldsymbol{\beta_{J-1}}$ for the general model LCPM which does not make an assumption of a common β.

For $j = 1, \ldots J-1$. Let the parameter vector for LCPM be $\Phi = (\alpha_1, \alpha_2, \ldots, \alpha_{J-1}, \boldsymbol{\beta}'_{\mathbf{1}}, \boldsymbol{\beta}'_{\mathbf{2}}, \ldots, \boldsymbol{\beta}'_{\boldsymbol{J-1}})'$ Denote the score statistic S_R:

$$S_R = \boldsymbol{S}'_{\text{LCPM}}(\hat{\boldsymbol{\Phi}}_0)(-\boldsymbol{H}_{\text{LCPM}})^{-1}\boldsymbol{S}_{\text{LCPM}}(\hat{\boldsymbol{\Phi}}_0),$$

where $\boldsymbol{S}_{\text{LCPM}}$ is the score function and $-\boldsymbol{H}_{\text{LCPM}}$ is the negative Hessian function for the LCPM, $\hat{\Phi}_0$ is the MLE from PPM. The score test in ordinal logistic regression[43] to test the proportional odds assumption provides that S_R is χ^2 distributed with $df = (J-2)p$. We anticipate a similar result for the log link here.

Recall the example in Section 4.2, which was about patient tonsil size and status as carriers of *Streptococcus pyogene*. The MLEs for a PPM are found in Table 3. The assessment of proportionality gives test statistic $S_R = 0.9055$, df = 1 and P value = 0.3413. Here there is no evidence against the proportionality assumption.

Recall the example in Section 4.3, which included a dataset for women with endometrial cancer. The MLEs for a PPM are found in Table 4. The assessment of proportionality gives test statistic $S_R = 12.38101$, df = 2 and P value = 0.00205. Here there is evidence against the proportionality assumption and the MLE from the LCPM seems more appropriate here.