Continuous Ordinal Regression for Analysis of Visual Analogue Scales: The **R** Package ordinalCont

Maurizio Manuguerra
Macquarie University

Gillian Z. Heller
Macquarie University

Jun Ma
Macquarie University

Abstract

This paper introduces the R package **ordinalCont**, which implements an ordinal regression framework for response variables which are recorded on a visual analogue scale (VAS). This scale is used when recording subjects' perception of an intangible quantity such as pain, anxiety or quality of life, and consists of a mark made on a linear scale. We implement continuous ordinal regression models for VAS as the appropriate method of analysis for such responses, and introduce smoothing terms and random effects in the linear predictor. The model parameters are estimated using constrained optimization of the penalized likelihood and the penalty parameters are automatically selected via maximization of their marginal likelihood. The estimation algorithm is shown to perform well, in a simulation study. Two examples of application are given: the first involves the analysis of pain outcomes in a clinical trial for laser treatment for chronic neck pain; the second is an analysis of quality of life outcomes in a clinical trial for chemotherapy for the treatment of breast cancer.

Keywords: continuous ordinal regression, visual analogue scale, maximum penalized likelihood, splines, R.

1. Introduction

Visual analogue scales (VAS) are used for measuring quantities which are intangible and difficult to measure on conventional scales, such as pain, anxiety and quality of life. These are generally used for self-rating. Subjects are given a linear scale of 100 mm and asked to put a mark where they perceive themselves. The scale has verbal anchor descriptors at each extreme, as depicted for the measurement of pain in Figure 1. The VAS reading is taken as the measurement from the left endpoint to the subject's mark, and is usually normalized to lie in the interval $[0, 1]$.

Figure 1: The visual analogue scale, with verbal descriptors for pain.

Statistical analysis of the VAS is controversial. While it is a bounded, continuous variable, several authors have argued that it is ordinal, rather than ratio in nature, and should be treated as such in analysis (for example Philip 1990; Wewers and Lowe 1990; Svensson 2000; Kersten, Küçükdeveci, and Tennant 2012). The issue is that the VAS cannot be considered a ratio measurement, or "linear", because for example a 1-cm difference in VAS scores at the lower end of the scale does not necessarily represent the same difference in the intangible outcome as a 1-cm difference at the upper end; and a doubling of VAS score may not translate to a doubling of e.g., the pain or anxiety. Whatever the extent of the "linear" portion in the VAS, it seems reasonable that some nonlinearity will be observed around the limits, where a higher density of observations, caused by perceptive states considered extreme or close to extreme, could be expected. The problem of non–interpretability of distances between measurements and the possibility of nonlinear behaviour, particularly at one or both extremes of the scale, is overcome by treating VAS measurements as ordinal rather than ratio data. We therefore refer to scales of this type as continuous ordinal.

We have developed a regression framework for continuous ordinal responses (Manuguerra and Heller 2010) and implemented it in the R package **ordinalCont**, which we present in this paper. The package, created in R (R Core Team 2020), is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=ordinalCont` (Manuguerra and Heller 2020). The relevance of continuous self-rating scales in the pain literature has been described in Heller, Manuguerra, and Chow (2016), where the frequent use of suboptimal methods in the analysis of VAS and the superior power of continuous ordinal regression analysis are discussed.

The remainder of this paper is structured as follows. Section 2 gives an overview of ordinal regression, in which the response variable is an ordered category. In Section 3 we develop a regression model based on responses such as VAS, which are measured on a bounded continuous scale but are at an ordinal level. Section 4 describes its implementation in the R package **ordinalCont**. Sections 6 and 7 give two examples of application, the first a pain response and the second quality of life.

2. Ordinal regression

Ordinal regression models (McCullagh 1980) are widely used for regression analysis of discrete ordinal responses Y within K ordered categories. The Y's are considered as coarse versions of an unobserved, continuous latent variable W, such that

$$Y = j \iff \alpha_{j-1} < W \le \alpha_j , \qquad j = 1, \ldots, K$$

where the α_j's are the correspondence on the latent variable scale of the category boundaries on the ordinal scale and $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_K = \infty$. This is depicted in Figure 2. Typically W is an intangible quantity such as pain, anxiety or quality of life and $y = 1, 2, \ldots, K$ codes for ordinal states such as "none", "mild", "moderate", "severe".
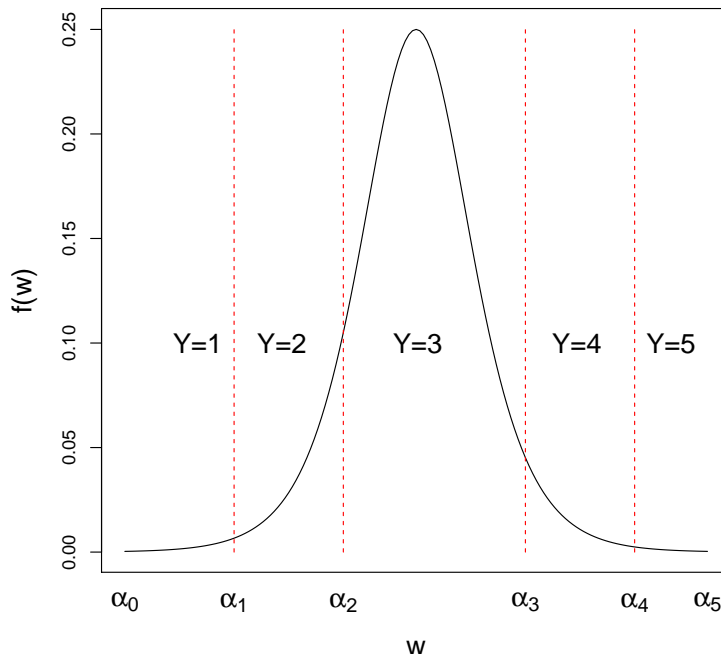
Figure 2: Correspondence between latent variable W and observed response Y, with $K = 5$.

The ordinal nature of Y is preserved by basing the regression model on the cumulative probabilities $\gamma_j = \mathsf{P}(Y \leq j)$. To relate the γ_j to covariates $x = (x_1, \ldots, x_p)^\top$ in the jth category, we write:

$$\gamma_j(x) = \mathsf{P}(Y \leq j \,|\, x) = \mathsf{P}(W \leq \alpha_j \,|\, x) .$$

The covariates are modelled on the latent scale:

$$W = -X^\top \beta + \epsilon , \tag{1}$$

where ϵ is a random error term, whose distribution we need to specify. A convenient choice for the distribution of ϵ is the standard logistic distribution, having cumulative distribution function (CDF) $\mathsf{F}(z) = \mathsf{P}(\epsilon \leq z) = 1/(1 + e^{-z})$ (Johnson, Kotz, and Balakrishnan 1995). Then

$$\gamma_j(x) = \mathsf{F}(\alpha_j + X^\top \beta) = \frac{1}{1 + e^{-(\alpha_j + X^\top \beta)}} .$$

Inverting this translates to the cumulative logistic model (also called the proportional odds model) for Y:

$$\ln \left(\frac{\gamma_j(x)}{1 - \gamma_j(x)} \right) = \alpha_j + X^\top \beta , \qquad j = 1, \ldots, K - 1 . \tag{2}$$

In the ordinal regression literature, the intercepts α_j are either ignored or used to characterize differences in category "size". In agreement with the latter approach, we argue that the α_js are worthy of careful modelling as they relate to an important cognitive aspect, i.e., how the perception of W changes at different levels. This behaviour can be of particular interest around the extremes of the scale, where steep curvature is indicative of a tendency for subjects to perceive and score their outcome at the extreme.

The ordinal regression model has developed to incorporate additive and non-linear functional forms of predictors (Hastie and Tibshirani 1987) and spline-based smoothers of predictors (Yee and Wild 1996). In Tutz (2003), a general framework to deal with semiparametrically structured models is given, with predictors that can contain parametric parts, additive parts with an unspecified functional form, and interactions. In particular global effects and category-specific effects are distinguished. In R, the standard ordinal regression is implemented in the package **ordinal** (Christensen 2019).

Typically VAS responses are analyzed by simply treating the VAS measurements as continuous responses in a normal model (for example Olivan-Blázquez *et al.* 2014); or by nonparametric methods such as the Mann-Whitney and Kruskal-Wallis tests (for example Mentula, Kalso, and Heikinheimo 2014); or by dichotomizing ("pain", "no pain") and using logistic regression (for example Choi *et al.* 2014); or by categorizing into a small number of levels and using model (2). Of these, the latter approach is most satisfactory, although it involves a loss of information. In a PubMed search we could only find one instance of it (Atrian, Abbaszadeh, Sarvieh, Sarafraz, and Jafarabadi 2013). The other approaches are less than optimal: use of the normal regression model has obvious distributional shortcomings; nonparametric methods are valid in that they are rank-based and therefore respect the ordinality of the measure; however, they are limited to comparisons between groups and cannot incorporate covariates or other more sophisticated regression features such as random effects.

We have proposed (Manuguerra and Heller 2010) a generalization of the standard ordinal model (2) which leads to an ordinal regression model for continuous scales. In this, the intercepts are modelled using parametric functions or B-splines. The original Bayesian estimation methodology has been updated (Manuguerra, Heller, and Ma 2017) to maximum penalized likelihood (MPL), which is faster and more stable. An extended version of this is described in Section 3. Similarly, Liu, Shepherd, Li, and Harrell Jr (2017) have used cumulative probability models for continuous outcomes. Their approach makes use of non-parametric maximum likelihood to estimate the parameters of the model and the intercepts and is then robust, but has the disadvantage of estimating a high number of parameters.

3. Regression model for continuous ordinal responses

3.1. The continuous ordinal regression model

Consider VAS measurements v which are sampled from a continuous response variable $V \in (0,1)$ [1], with density $f_v(v)$ and CDF $\gamma(v)$. The continuous ordinal response variable V reflects the subjective perception of the underlying continuous latent variable W defined on the real line. The dependence between V and W is modelled by a smooth one-to-one function $g : (0,1) \mapsto \mathbb{R}$ that maps v on the VAS scale to $w = g(v)$ on the latent scale. This mapping is the link between the recorded perception of the latent variable and an underlying metric. As for the standard ordinal model, the covariates X are modelled on the latent scale. Defining the regressor $h^\star(X)$ and assuming $W = -h^\star(X) + \epsilon$, the cumulative distribution function for the score V can then be written as:

$$\gamma(v|X) = \mathsf{P}(V \leq v|X) = \mathsf{P}(W \leq g(v)|X) = \mathsf{F}(g(v) + h^\star(X)) = \mathsf{F}(h(v,X)) \qquad (3)$$

---

[1]For mathematical convenience and without loss of generality, we assume in what follows that v is scaled to lie in $(0,1)$.

From Equation 3 it is clear that the ordinal regression model is a transformational model (Möst and Hothorn 2015; Hothorn, Möst, and Bühlmann 2018) for the conditional distribution of V, which depends on the distribution function of ϵ, $\mathsf{F}(\cdot)$, and the regressor $h(v, X) = g(v) + h^\star(X)$.

The function $g(v)$ is the continuous analog of the discrete intercepts α_j in model (2), and its shape is informative of the change in perception of the latent variable at different levels (as are the α_j). The systematic component $h^\star(X)$ may incorporate fixed effects, random effects and smoothing terms. Random effects are useful not only for modelling clustered or longitudinal data, as in the examples in Sections 6 and 7, but also for modelling individual variation due to subjective perception of the latent variable.

Differentiating Equation 3, we can write the density of V as:

$$\mathsf{f}_v(v \mid X) = \frac{\partial \gamma(v \mid X)}{\partial v} = \dot{\mathsf{F}}(h(v, X)) \, \frac{\partial h(v, X)}{\partial v} = \dot{\mathsf{F}}(h(v, X)) \, \dot{g}(v) \tag{4}$$

where $\dot{\mathsf{F}}(h) = \frac{\partial \mathsf{F}(h)}{\partial h}$, $\frac{\partial h(v,X)}{\partial v} = \frac{\partial (g(v)+h^\star(X))}{\partial v} = \frac{\partial g(v)}{\partial v}$ and $\dot{g}(v) = \frac{\partial g(v)}{\partial v}$.

The log-likelihood is obtained taking the logarithm of Equation 4. Re-ordering the terms, for subject i it can be written in the general form:

$$\ell_{0i} = \log\left(\dot{g}(v_i)\right) + \log\left(\dot{\mathsf{F}}(h(v_i, X_i))\right) \tag{5}$$

A variety of distributions can be used for $\mathsf{F}(\cdot)$, and some of the most popular ones (logit, probit, loglog, cloglog, Cauchit) are implemented in the **ordinalCont** package. We can comment that if the cumulative distribution is $\mathsf{F}(h) = \frac{e^h}{1+e^h}$ (logistic model), the regressor $h(v, X)$ can be interpreted as the log-odds and the systematic component $h^\star(X)$ as the log-odds ratio to an observation with $h^\star(X) = 0$, regardless of the value of v. When $\mathsf{F}(h) = 1 - e^{-e^h}$ (cloglog link function) and the score v is the time to event, the model can be used for survival analysis and the g function is the log-cumulative hazard at baseline. From Equations 3 and 4, we can derive the general forms of the survival function $\mathsf{S}(v|X) = 1 - \mathsf{F}(h(v, X))$, the cumulative hazard function $\Lambda(v|X) = -\log(1 - \mathsf{F}(h(v, X)))$ and the hazard function $\lambda(v|X) = \frac{\dot{\mathsf{F}}(h)\dot{g}(v)}{1-\mathsf{F}(h(v,X))}$.

3.2. The g function

In order to specify model (5), we need to define the form of the function $g(v)$. It has to be differentiable, monotonic increasing and "flexible enough" to capture the nonlinear behaviour of the ordinal measure. In the contest of the ordinal regression, where the recorded scores are always bounded and can be normalised to the interval $[0, 1]$, any function which maps $(0, 1)$ to $\mathbb{R}$ could be appropriate (see next paragraph for how to transform data from a closed to an open interval). The choice of such a function does not decrease the generality of the formulation of the problem and its applicability to other models (e.g., transformational models which are not ordinal models), as real data are always bounded in an interval and can then be normalised. The function $g(v)$ has previously been expressed using a parametric approach, as in Manuguerra and Heller (2010), or non-parametrically, as in Liu *et al.* (2017). In this study we have implemented an approach based on monotonic I-splines Ψ_u (Ramsay 1988). We can then write $g(v) = \sum_{u=1}^{p_\theta} \theta_u \Psi_u(v)$, where p_θ is the number of basis functions. This specific formulation allows for more flexibility than the parametric approach, not assuming the shape of the function $g(v)$; makes it easy to satisfy the monotonicity constraint, by imposing $\theta_u \geq 0, \forall u$; provide the first two derivatives in easy-to-express forms, i.e., $\dot{g}(v) = \sum_{u=1}^{p_\theta} \theta_u \dot{\Psi}_u(v)$

and $\ddot{g}(v) = \sum_{u=1}^{p_\theta} \theta_u \ddot{\Psi}_u(v)$, where $\dot{\Psi}_u(v)$ and $\ddot{\Psi}_u(v)$ can be efficiently computed using the properties of I-splines and M-Splines.

Scale endpoints

A problem in extending the ordinal regression model to continuous scales is that the VAS is a closed scale, i.e., it contains the extremes 0 and 1. As $\lim_{v\to 0} g(v) = -\infty$ and $\lim_{v\to 1} g(v) = \infty$, the presence of either 0 or 1 results in the model not being identifiable. Both extremes are valid recordings, in particular $v = 0$, reflecting an absence of the latent variable (e.g., pain or anxiety). In order to deal with this numerically, where necessary we use the scaling proposed by Smithson and Verkuilen (2006). Assuming raw VAS observations v^*, use v defined as

$$v = \frac{(n-1)v^* + 0.5}{n} ,$$

where n is the sample size. This scales $v^* \in [0, 1]$ to $v \in (0, 1)$.

3.3. Covariates

In the current formulation, the roles of the observed score v and the covariates X in the regressor $h(v, X)$ are separated in the two terms $g(v)$ and $h^\star(X)$. Assuming p fixed effects, r individual effects and s smoothing terms, the latter can be written as:

$$h^\star(X) = \sum_{m=1}^{t} X_m \eta_m = U\beta + \sum_{j=1}^{r} Z_j b_j + \sum_{l=1}^{s} \Phi_l(Y_l)\vartheta_l \tag{6}$$

where U is a $n \times p$ design matrix for the fixed effects, Z_j is the design matrix for the j-th individual effect and Φ_l is the matrix containing the basis functions (B-splines) evaluated at the values of the covariate Y_l. Here η is the set of parameters $\{\eta_1^\top, \ldots, \eta_t^\top\}^\top = \{\beta^\top, b^\top, \vartheta^\top\}^\top$, with $\beta^\top = \{\beta_1, \ldots, \beta_p\}$, $b^\top = \{b_1^\top, \ldots, b_r^\top\}$, $\vartheta^\top = \{\vartheta_1^\top, \ldots, \vartheta_s^\top\}$, $t = p + r + s$ and X_m is the mth element in $\{U_1, \ldots, U_p, Z_1, \ldots Z_r, \Phi_1, \ldots \Phi_s\}$.

3.4. Penalized likelihood estimation

B-splines, I-splines and individual effects in the model can lead to overfitting. We avoid it using penalty functions, i.e., including additional constraints on ϑ and θ, the parameters that define the smoothing terms and the g function. Our experience (see also Ruppert, Wand, and Carroll 2003, page 66) reveals that a penalty function can stabilize the estimation process and makes the selection of number and location of knots less important.

The penalized log-likelihood can then be written as:

$$\begin{aligned}
\ell_p &= \sum_{i=1}^{n} \ell_{0i} - \lambda_g J_g(\theta) - \sum_{j=1}^{r} \lambda_{b_j} J(b_j) - \sum_{l=1}^{s} \lambda_{\vartheta_l} J(\vartheta_l) \\
&= \sum_{i=1}^{n} \ell_{0i} - \lambda_g J_g(\theta) - \sum_{m=1}^{t} \lambda_m J_m(\eta_m)
\end{aligned} \tag{7}$$

where the J's are penalty functions, such as roughness penalties, and the first p terms in the last sum account for penalty terms relative to the fixed effects, and are zero. The roughness

penalties can be written in general as $J(\theta) = \theta^\top R\theta$ where the R terms are square matrices with elements given by $R_{jk} = \int \ddot{\Psi}_j(y)\ddot{\Psi}_k(y)\,dy$. For the individual effect terms, R is an identity matrix and the penalty term is $J(b) = b^\top b$, which implies normally distributed random effects b with variance $\sigma_b^2 = \frac{1}{2\lambda_b}$.

The model parameters we want to estimate are $\eta = \{\eta_1^\top, \ldots, \eta_t^\top\}^\top$ and $\theta = \{\theta_1, \ldots, \theta_m\}^\top$. Their constrained MPL estimates are then defined by

$$(\hat{\eta}, \hat{\theta}) = \arg\max_{\eta,\theta}\left\{\ell_p = \sum_{i=1}^{n} \ell_{0i} - \lambda_g J_g(\theta) - \sum_{m=1}^{t} \lambda_m J_m(\eta_m)\right\}. \tag{8}$$

3.5. Estimation of model parameters

The estimation procedure iterates through two steps repeated until convergence. First, given the current values of the λ's, the parameters that define the g function and the predictor, θ and η, are estimated, then given the current value of θ and η the λ's are calculated according to the method detailed in Section 3.6. The first estimation of the parameters θ and η is done using maximum likelihood, i.e., with the starting value $\lambda^0 = 0$. When the number of observations is large compared to the number of parameters, this choice does not create convergence problems, but the user should be careful in the opposite case. We comment that this is a constrained optimization as $\theta \geq 0$, and that $g(v)$ and $h(v, X)$ are estimated jointly, as they are updated at every step.

The Karush-Kuhn-Tucker (KKT) necessary conditions for the constrained MPL estimation of θ and η are:

$$\frac{\partial \ell_p}{\partial \eta_m} = 0 \qquad \text{for any } m \tag{9}$$

$$\frac{\partial \ell_p}{\partial \theta_u} = \begin{cases} 0 & \text{if } \theta_u > 0 \\ < 0 & \text{if } \theta_u = 0. \end{cases} \tag{10}$$

Equations 9 and 10 can be solved iteratively, with the unconstrained parameters η estimated using a Newton method and the positively constrained parameters θ estimated using the multiplicative iterative (MI) algorithm (Ma 2010).

Given the estimated values of η and θ at iteration k, the values of each $\eta_m \in \eta$ at iteration $k+1$ are obtained with the Newton algorithm:

$$\eta_m^{(k+1)} = \eta_m^{(k)} + \omega^{(k)}\left(\frac{\partial^2 \ell_p}{\partial \eta_m \partial \eta_m^\top}(\eta^{(k)}, \theta^{(k)})\right)^{-1}\left(\frac{\partial \ell_p}{\partial \eta_m}(\eta^{(k)}, \theta^{(k)})\right)$$

where $\omega^{(k)} \in (0, 1]$ is a line search step size that guarantees that $\ell_p(\eta^{(k+1)}, \theta^{(k)}) \geq \ell_p(\eta^{(k)}, \theta^{(k)})$. The first and the second derivative of the penalized log-likelihood (7) with respect to η are:

$$\nabla_\eta \ell_p = \nabla_\eta h^\top \left(\frac{\ddot{F}(h)}{\dot{F}(h)}\right) - 2\lambda R\eta$$

and

$$\nabla_\eta^2 \ell_p = \nabla_\eta h^\top \ L \ \nabla_\eta h - 2\lambda R$$

where $\nabla_\eta h^\top$ is the transpose of the $n \times t$ matrix with elements $\frac{\partial h_i(v,\eta)}{\partial \eta_m}$, with η_m the m-th element of the vector η, L is the $n \times n$ diagonal matrix with $\text{diag}(L) = \frac{\ddot{\mathsf{F}}(h)\dot{\mathsf{F}}(h) - \ddot{\mathsf{F}}(h)^2}{\dot{\mathsf{F}}(h)^2}$, R is a block diagonal matrix with R_m matrices on the diagonal and $\lambda = \lambda_1, \ldots, \lambda_t$.

Given the estimated values of η at iteration $k + 1$ and θ at iteration k, the values of θ at iteration $k+1$ are obtained with the MI algorithm (Ma, 2010). The MI algorithm can handle a large number of non-negatively constrained parameters as it is very efficient to implement, involving only the first derivative of the penalized likelihood function. When $\theta > 0$, by the second KKT condition (10):

$$\nabla_\theta \ell_p = \Psi^\top \left(\frac{\ddot{\mathsf{F}}(h)}{\dot{\mathsf{F}}(h)} \right) + \Psi'^\top E - 2\lambda_\theta R_\theta \theta = 0$$

where E is a diagonal matrix with $\text{diag}(E) = \frac{1}{g'(v)}$ and 0 otherwise.

Following Ma (2010), it is straightforward to derive the updating algorithm:

$$\theta^{(k+1)} = \theta^{(k)} + \omega_\theta^{(k)} s^{(k)} \nabla_\theta \ell_p(\eta^{(k+1)}, \theta^{(k)}),$$

where a line search of size $\omega^{(k)} \in (0,1]$ is introduced to guarantee that $\ell_p(\eta^{(k+1)}, \theta^{(k+1)}) \geq \ell_p(\eta^{(k+1)}, \theta^{(k)})$, $s^{(k)} = -\frac{\theta^{(k)}}{\left[\nabla_\theta \ell_p(\eta^{(k+1)}, \theta^{(k)}) \right]^-_{+\tau}}$, $[x]^- = x$ if $x \leq 0$, and 0 otherwise, and τ is a small constant introduced to avoid computational instability.

3.6. Automatic smoothing parameter estimation

Fixing the values of η to the most current estimate, the optimal smoothing parameters are the roots of the partial derivatives with respect to each λ_m of the marginal posterior computed integrating out θ and η from (7). The log-posterior can be written as:

$$\Phi = \sum_{i=1}^n \ell_{0i} - \left(\frac{p_\theta}{2} \log(\sigma_\theta^2) + \frac{1}{2\sigma_\theta^2} \theta^\top R\theta \right) - \sum_{m=1}^t \left(\frac{p_m}{2} \log(\sigma_m^2) + \lambda_m \eta_m^\top X_m \eta_m \right)$$

where p_m is the number of parameters estimated for the m-th smoother or random effect. From here the marginal log-posterior is obtained by integrating out θ and η:

$$\Phi_{\text{marg}} = \log \int \int \exp \left(\Phi(\theta, \eta; \sigma_\theta^2, \sigma_\eta^2) \right) d\theta \, d\eta$$

Exact solution to this integral is infeasible, but it can be approximated following Laplace (1774):

$$\Phi_{\text{marg}} = -\frac{p_\theta}{2} \log(\sigma_\theta^2) - \sum_{m=1}^t \frac{p_m}{2} \log(\sigma_m^2) + \ell_p(\hat{\theta}, \hat{\eta}; \sigma_\theta^2, \sigma_\eta^2) - \frac{1}{2} \log \left| -Q(\hat{\theta}, \hat{\eta}; \sigma_\theta^2, \sigma_\eta^2) \right|$$

where $Q(\theta, \eta; \sigma_\theta^2, \sigma_\eta^2) = \nabla_{\theta\eta}^2 \ell_p(\theta, \eta; \sigma_\theta^2, \sigma_\eta^2)$. Solving

$$\frac{\partial \Phi_{\text{marg}}}{\partial \sigma_\theta^2} = 0$$

$$\frac{\partial \Phi_{\text{marg}}}{\partial \sigma_{\eta_m}^2} = 0$$

for σ_θ^2 and $\sigma_{\eta_m}^2$ we obtain the desired smoothing parameters estimates:

$$\sigma_\theta^2 = \frac{1}{2\lambda_\theta} = \frac{\theta^\top R_\theta \theta}{p_\theta - \mathrm{tr}\left(Q^{-1}Q_\theta\right)}$$

$$\sigma_m^2 = \frac{1}{2\lambda_m} = \frac{\eta_m^\top R_m \eta_m}{p_m - \mathrm{tr}\left(Q^{-1}Q_m\right)}$$

where $Q_\theta = \lambda_\theta \nabla_{\theta\eta}^2 (\theta R\theta)$, $Q_m = \lambda_m \nabla_{\theta\eta}^2 (\eta_m A_m \eta_m)$ and $\nabla_{\theta\eta}^2 = \frac{\partial^2}{\partial\theta\partial\eta}$.

3.7. Quantile residuals

The estimated CDF of V, commonly called probability integral transform, is

$$\hat{\mathsf{F}}(v \mid X) = \mathsf{F}\left(\hat{h}(v, X)\right) = \mathsf{F}\left(\hat{g}(v) + \hat{h}^\star(X)\right)$$

where $\hat{g}(v)$ and $\hat{h}^\star(X)$ are calculated at the estimated values of θ and η. Quantile residuals (Dunn and Smyth 1996) are computed at observed v_i as

$$q_i = \Phi^{-1}\left(\hat{\mathsf{F}}(v_i \mid X_i)\right) , \qquad i = 1, \dots, n ,$$

where $\Phi^{-1}(\cdot)$ is the inverse standard normal CDF. If model assumptions are satisfied and there is sufficient diversity in the observations (v_i, X_i), then the residuals q_i should be normally distributed. On the other hand, if observations are highly discretized, for example at value $v = 0$ with limited covariate values, the q_i could deviate from normality. In this case, we recommend the use of randomized quantile residuals (Dunn and Smyth 1996).

4. Package overview

The **ordinalCont** R package implements the continuous ordinal regression models, using maximum penalized likelihood estimation. The main function is `ocm` for both fixed- and mixed-effect models. Model equations are specified with syntax as in, for example, function `lm` and the package **lme4** (Bates, Mächler, Bolker, and Walker 2015). The output of the `ocm` function is an object of class 'ocm', and several methods are provided for it, including `plot`, `print`, `summary`, `extractAIC`, `logLik`, `predict` and `anova`. The `plot` method displays the following graphical summaries:

- a plot of the estimated g function with 95% confidence intervals;

- the histogram and normal Q-Q plot of the quantile residuals;

- plots of the smooth terms (B-splines) for covariates x, specified as `s(x)` in the formula (if included in the model).

The `predict` method provides, for given covariate values X^* and parameter estimates $\hat{\eta}$ and $\hat{\theta}$, the predicted values for v, the conditional density function of v, the conditional distribution function, the conditional quantile function, the linear regressor, the exponential of the linear

regressor, the conditional hazard functions, the conditional cumulative hazard function and the conditional survival function.

5. Simulation

To test the accuracy and precision of the model parameter estimates, we have generated random data according to the model:

$$h(\eta_i, v_i) = g(v_i) + \beta_1 a_1 + \beta_2 a_2 + f(a_3) + b_i,$$

where i is an index identifying the subjects in the sample, $\beta_1 = 1.7$, $\beta_2 = -0.4$, $f(a_3) = 2 - 2a_3 + 0.1a_3^2 + \sin(a_3)$, $b_i \sim N(0, \sigma_b = 4)$ are random effects and the g function is the generalized logistic function (Richards 1959):

$$g(v) = M + \frac{1}{B} \log \left(\frac{Tv^\top}{1 - v^\top} \right), \qquad 0 < v < 1$$

with parameters chosen to make it fairly asymmetric ($M = 0.2$, $B = 0.3$, $T = 2$). The covariates have been sampled from arbitrary normal and uniform distributions, specifically $a_1 \sim \mathcal{N}(\mu = 10, \sigma^2 = 9)$, $a_2 \sim U(3, 5)$, $a_3 \sim U(0, 20)$.

5.1. Results

In the simulations we have used four sample sizes, $n = 100, 200, 500, 1000$, and for each sample size we have generated and analysed 1000 samples. In Table 1 we report the means and the standard deviations of the 1000 estimates, while in Figures 3 and 4 the estimates of the smoother and the g function are shown together with the true values of the functions for a typical simulation.

As expected with penalized likelihood regressions, increasing the sample size has the effect of decreasing the bias and the variance of the parameter estimates. It is worth noting that both the variance of the random effects and the nonlinear effect of a_3 are well estimated for all the sample sizes, $n = 100$ included. In the latter case, despite the limited number of observations, the estimates are still good indicators of the strength and direction of the effects.

| | β_1 (1.7) | | β_2 (−0.4) | | σ_b (5) | | penalized logLik | | edf | |
n	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
100	1.489	0.200	−0.361	0.346	4.264	1.181	199.08	27.22	29.52	2.18
200	1.582	0.161	−0.378	0.235	4.539	1.181	406.01	52.83	33.99	1.86
500	1.652	0.104	−0.395	0.136	4.722	1.204	1036.37	105.69	38.37	1.27
1000	1.673	0.063	−0.390	0.104	4.784	1.135	2072.27	189.17	40.59	0.72

Table 1: Simulation results: means and standard deviations of the estimates, penalized log-likelihood values, and equivalent degrees of freedom (edf).
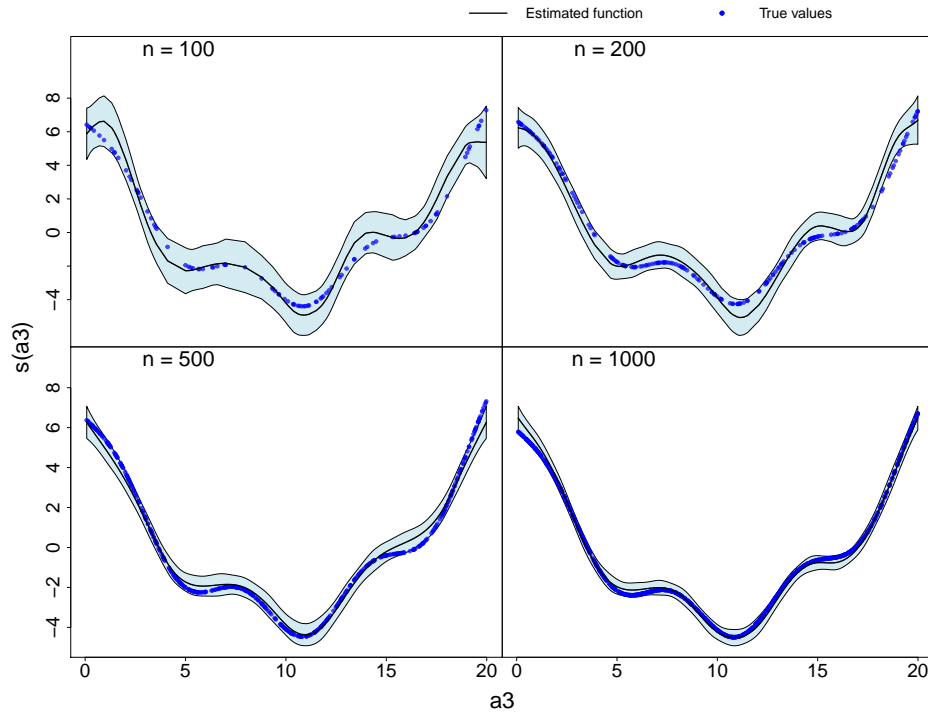
Figure 3: Typical estimated smoothing function $\hat{f}(a_3)$ with confidence bars, by sample size. The blue dots in the plot are the true values of $f(a_3)$.
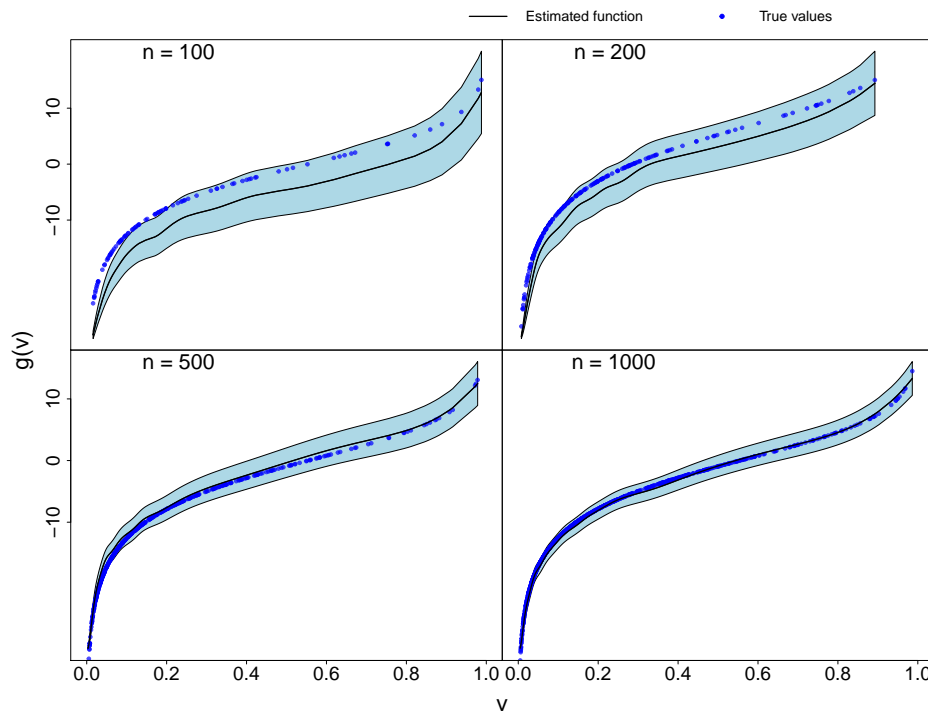

Figure 4: Typical estimated g function with confidence bars by sample size. The blue dots in the plot are the values of $g(v)$ as in the model.

6. Application 1: Chronic neck pain study

In this example, we illustrate the use of the continuous ordinal model when treatment effects, interaction effects and random effects are included in the model.

This study (Chow, Heller, and Barnsley 2006) is a randomized controlled trial on the efficacy of low-level laser therapy (LLLT) in the treatment of chronic neck pain. Patients with chronic neck pain were randomized to receive 14 LLLT treatments over 7 weeks, with either active or placebo laser (`laser = 1`, `laser = 2` respectively). The primary outcome measure of the study was the VAS, marked by the patients at baseline, at the end of the course of treatment and one month after completion of the treatment. We analyzed data from $n = 84$ subjects measured at the three time points. The VAS scores are in $[0, 1]$, and the `ocm` function automatically scales them to $(0, 1)$ for the purpose of analysis.

The data layout is (first three cases):

```
R> neck_pain[1:9, ]
```

```
   id  vas laser time
1   1 0.91     1    1
2   1 0.31     1    2
3   1 0.51     1    3
4   2 0.12     2    1
5   2 0.30     2    2
6   2 0.30     2    3
7   3 0.18     2    1
8   3 0.18     2    2
9   3 0.18     2    3
```

To illustrate the `ocm` function, we model just a treatment effect at the end of the treatment period (`time = 2`). The continuous ordinal model is:

$$\log\left(\frac{\gamma(v_i)}{1 - \gamma(v_i)}\right) = g(v_i) + x_{1i}\beta_1 \qquad i = 1, \ldots, n \ , \tag{11}$$

where x_1 is an indicator variable for treatment.

```
R> lasermodel0 <- ocm(vas ~ laser, data = subset(neck_pain, time == 2))
R> summary(lasermodel0)


Call:
ocm(formula = vas ~ laser, data = subset(neck_pain, time == 2))


Coefficients:
       Estimate  StdErr t.value  p.value
laser1  1.41659 0.40572  3.4915 0.004489 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
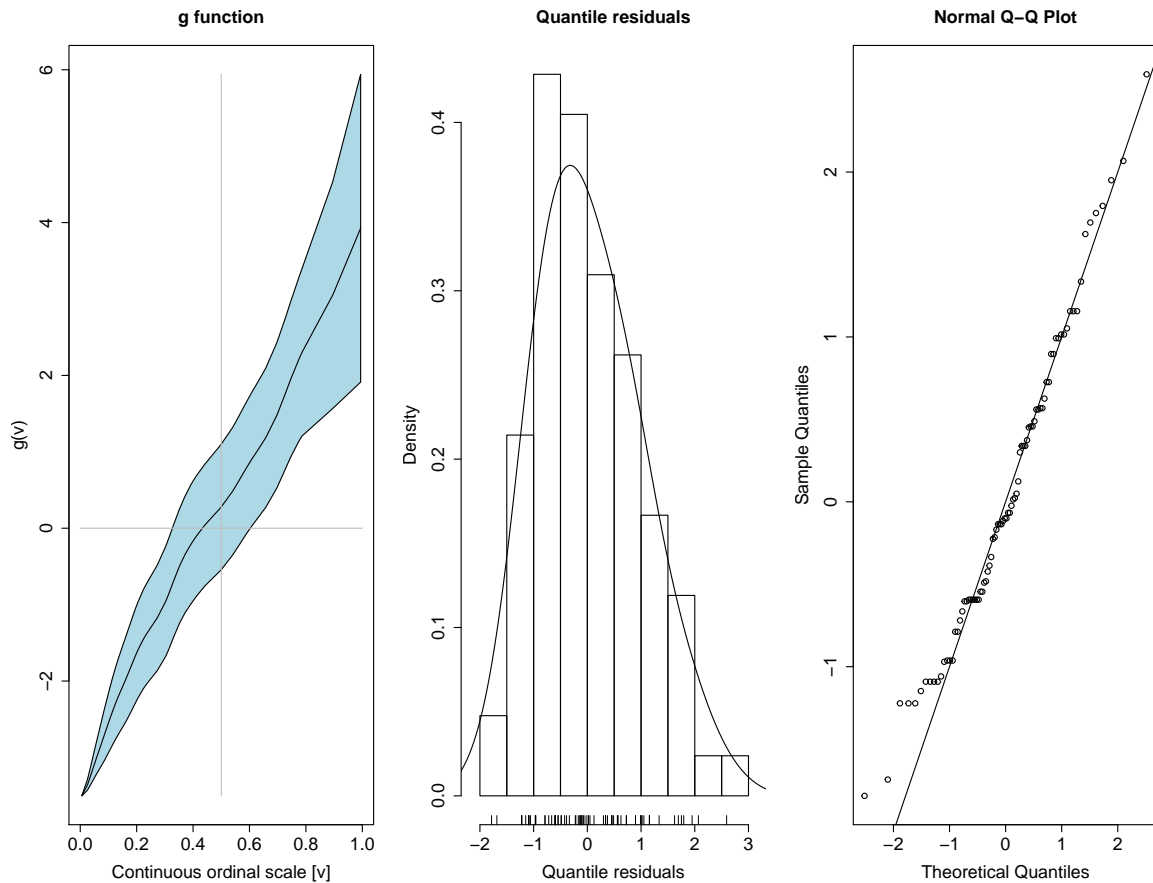
Figure 5: Estimated g function, histogram of quantile residuals and quantile residual normal Q-Q plot for fitted model `lasermodel0`.

We obtain a significant treatment effect ($p = 0.004$), with $\hat{\beta}_1 = 1.417$. Since e^{β_1} is the effect of the active laser treatment on the odds of scoring $V \leq v$, positive β_1 signifies lower VAS scores due to laser, i.e., pain relief.

To check the model fit, we look at the histogram of the quantile residuals and the quantile residual normal Q-Q plot:

```
R> plot(lasermodel0)
```

The plots (Figure 5) suggest an adequate model fit, even if with a slightly skewed distribution of the quantile residuals, confirmed by the Q-Q plot.

To model the VAS scores at the three time points, we use a random effect for patient. The continuous ordinal model is

$$\log\left(\frac{\gamma(v_{it})}{1 - \gamma(v_{it})}\right) = g(v_{it}) + x_{1i}\beta_1 + x_{2t}\beta_2 + x_{3t}\beta_3 + x_{1i}x_{2t}\beta_{12} + x_{1i}x_{3t}\beta_{13} + b_i$$

$$i = 1, \ldots, n; \ t = 1, 2, 3 \ ,$$

where x_2 and x_3 are indicator variables for times $t = 2$ and $t = 3$ respectively, and within-patient correlation is modelled using the random effect $b_i \sim \mathcal{N}(0, \sigma^2)$ for patient i. The function `ocm` is used, with the term `(1|id)` specifying the random intercept for patient (`id`):

```
R> lasermodel1 <- ocm(vas ~ laser * time + (1 | id), data = neck_pain)
R> summary(lasermodel1)


Call:
ocm(formula = vas ~ laser * time + (1 | id), data = neck_pain)

Random effects:
        Name Variance Std.Dev.
 Intercept|id    6.07    2.464

Coefficients:
            Estimate   StdErr t.value   p.value
laser1       -2.13735  0.63884 -3.3457  0.001171 **
time2        -0.59288  0.35276 -1.6807  0.096065 .
time3        -0.18980  0.34430 -0.5513  0.582729
laser1:time2  4.45469  0.55559  8.0180 2.539e-12 ***
laser1:time3  3.37869  0.51837  6.5179 3.258e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significant treatment benefits are demonstrated at times 2 and 3, as evidenced by the significance of the interaction terms `laser1:time2` ($\hat{\beta}_{12}$) and `laser1:time3` ($\hat{\beta}_{13}$), with an attenuation of the positive effect at time 3. Note that there was an imbalance between treatment groups at baseline, with the active treatment group having worse pain ($\hat{\beta}_1 = -2.14, p = 0.001$).

The random effects (b_i) are sampled from a normal distribution with significant variance $\hat{\sigma}^2 \approx 6.1$, signifying a persistent subject effect. It is particularly important to incorporate this effect because of the individual and subjective nature of the measurements.

We use the `predict` method to compare the observed and predicted values of the VAS score.

```
R> plot(lasermodel1$v, predict(lasermodel1), xlab = "observed score",
+    ylab = "predicted score", xlim = c(0, 1), ylim = c(0, 1))
R> lines(c(0, 1), c(0, 1))
```

The results are shown in Figure 6. The reader could notice that attenuation is present in the plot, as predicted scores are higher than observed scores for low values of v, and vice-versa. We can comment that, based on our experience with simulated and real data sets, this feature is not commonly observed, even when a high proportion of the observed score v takes values 0 or 1.

In order to appreciate the estimated effect of the treatment on the score v, and under the assumption of individual effect $b = 0$, we can calculate the conditional density distributions $f(v|\text{treatment} = \text{"laser"}, \text{b=0})$ and $f(v|\text{treatment="placebo"}, \text{b=0})$, and the conditional cumulative distributions $F(v|\text{treatment="laser"}, \text{b=0})$ and $F(v|\text{treatment="placebo"}, \text{b=0})$.

```
R> n <- nrow(neck_pain)
R> vas <- seq(0.01, 0.99, len = n)
R> laser <- placebo <- neck_pain$laser
```
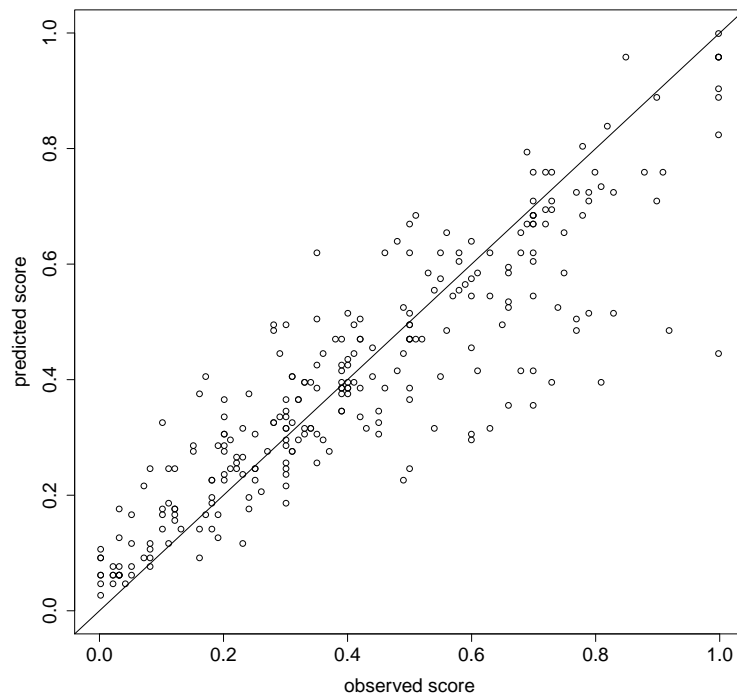
Figure 6: Observed and predicted scores for fitted model `lasermodel1`.

```
R> laser[laser != 1] <- 1
R> placebo[placebo != 2] <- 2
R> time <- neck_pain$time
R> time[time != 3] <- 3
R> plot(vas, predict(lasermodel1,
+     newdata = data.frame(vas = vas, laser = laser, time = time),
+     type = "density"),
+     t = "l", col = "red", xlab = "v", ylab = "f(v)", ylim = c(0,4))
R> lines(vas, predict(lasermodel1,
+         newdata = data.frame(vas = vas, laser = placebo, time = time),
+         type = "density"))
R> legend("topright", c("Active laser", "Placebo"),
+         col = c("red", "black"), lty = c(1,1))
```

The results, shown in Figure 7, illustrate how subjects treated with active laser (`laser = 1`) experience lower values of VAS at `time = 3` than those treated with placebo (`laser = 2`).

7. Application 2: Breast cancer study

In this example, we demonstrate the use of a smooth term for one of the covariates and of the `predict.ocm` method.

Metastatic breast cancer is the most common cause of cancer death among Australian women. The ANZ0001 trial is a randomized trial with three chemotherapy treatment arms ($n =$
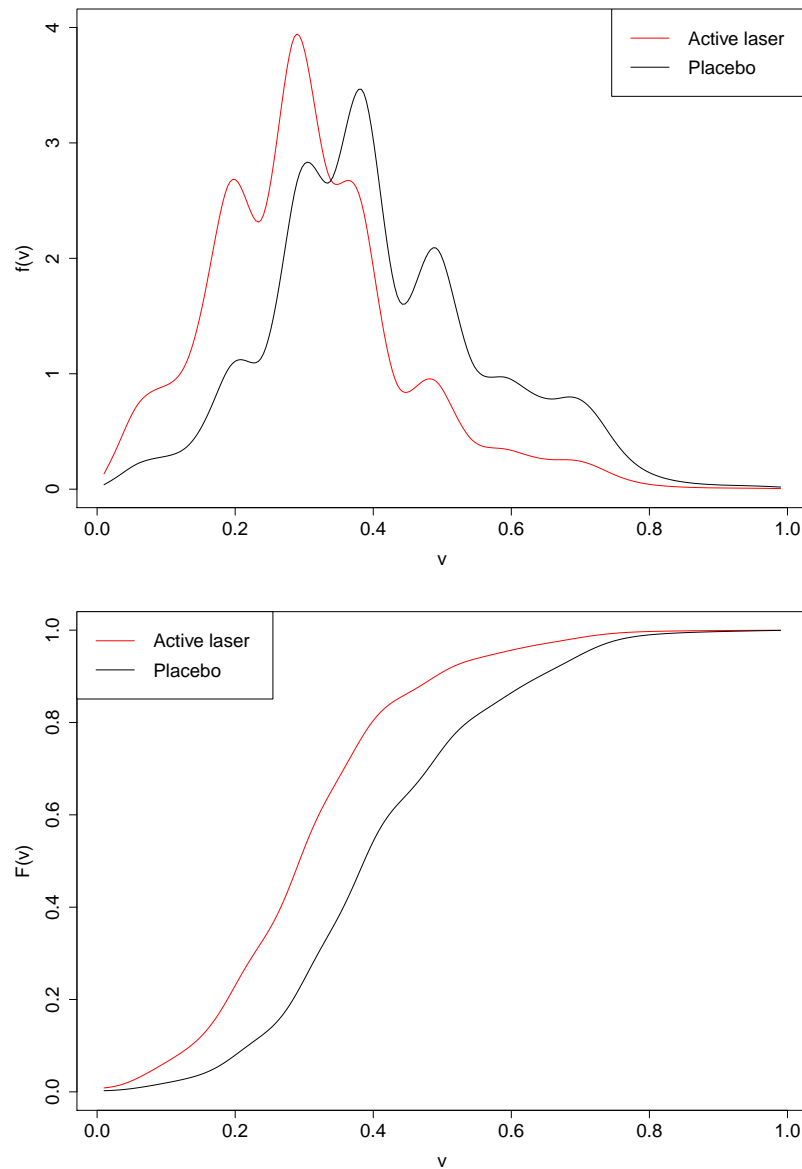
Figure 7: Estimated density distributions of the score v conditional on treatment and random effects for fitted model `lasermodel1`, at `time = 3`.

292 patients with complete quality of life measurements) concluded in 2005 (Stockler *et al.* 2011). Health-related quality of life is assessed at each chemotherapy treatment cycle. The treatments Intermittent Capecitabine (IC) and Continuous Capecitabine (CC) are compared with the standard combination treatment CMF, each with its own protocol. There is no maximum duration of treatment, but it is interrupted on disease progression, or when patient intolerance or unacceptable toxicity are recorded.

In this analysis we aim to verify which treatment has a better impact on quality of life, and in particular how this impact changes over chemotherapy cycle. Various aspects of quality

Cycle	Treatment		
	IC	CC	CMF
0	85	85	90
5	61	56	51
10	34	38	10
15	19	21	3
20	13	11	2
25	7	6	0
45	1	3	0
60	0	1	0

Table 2: Number of patients at different cycle numbers.

of life are assessed using the VAS[2]. The number of remaining patients in selected cycles, by treatment arm, are shown in Table 2. Note that no patients on standard CMF treatment progressed beyond cycle 20, and for the other two treatments the data are sparse beyond this point. We should comment that the software does not natively accommodate censoring at the moment. This includes informative censoring, e.g., cases in which subjects quit the study for causes correlated with the covariates (death, relapses, etc.). In these cases, the results could be biased.

We used the overall quality of life VAS scale as dependent variable. Among the several covariates available, only treatment and chemotherapy cycle number were found to be significant.

We fitted eight models for the overall quality of life v_{ij} for patient i at chemotherapy cycle j:

Model 1: $\sim g(v_{ij}) + t_i + j$
Model 2: $\sim g(v_{ij}) + t_i + j + b_i$
Model 3: $\sim g(v_{ij}) + t_i * j$
Model 4: $\sim g(v_{ij}) + t_i * j + b_i$
Model 5: $\sim g(v_{ij}) + t_i + s(j)$
Model 6: $\sim g(v_{ij}) + t_i + s(j) + b_i$
Model 7: $\sim g(v_{ij}) + t_i * s(j)$
Model 8: $\sim g(v_{ij}) + t_i * s(j) + b_i$

where t_i is the treatment arm of patient i, $s(j)$ is a smooth term that depends on the cycle number j, $t_i * s(j) = t_i + s(j) + s(j|t_i)$, $s(j|t_i)$ is a smooth term that depends on cycle number j conditional on the treatment arm, b_i are random effects sampled from $\mathcal{N}(0, \sigma^2)$ and the error on the latent scale follows the logistic distribution.

```
R> model1  <- ocm(overall  ~ treatment + cycleno,
+    data = ANZ0001, scale = c(0, 100))
R> model2  <- ocm(overall  ~ treatment + cycleno + (1 | randno),
+    data = ANZ0001, scale = c(0, 100))
R> model3  <- ocm(overall  ~ treatment * cycleno,
+    data = ANZ0001, scale = c(0, 100))
R> model4  <- ocm(overall  ~ treatment * cycleno + (1 | randno),
+    data = ANZ0001, scale = c(0, 100))
```

---

[2]The term used in the quality of life literature for VAS is Linear Analog Self-Assessment (LASA).

```
R> model5  <- ocm(overall  ~ treatment + s(cycleno),
+    data = ANZ0001, scale = c(0, 100))
R> model6  <- ocm(overall  ~ treatment + s(cycleno) + (1 | randno),
+    data = ANZ0001, scale = c(0, 100))
R> model7  <- ocm(overall  ~ treatment * s(cycleno),
+    data = ANZ0001, scale = c(0, 100))
R> model8  <- ocm(overall  ~ treatment * s(cycleno) + (1 | randno),
+    data = ANZ0001, scale = c(0, 100))
```

We choose the best model comparing the AICs of the models:

```
R> anova(model1, model2, model3, model4, model5, model6, model7, model8)
```

```
[Penalized] likelihood ratio tests of ordinal regression models for
continuous scales:

        formula                                          link
model1 overall ~ treatment + cycleno                     logit
model3 overall ~ treatment * cycleno                     logit
model5 overall ~ treatment + s(cycleno)                  logit
model7 overall ~ treatment * s(cycleno)                  logit
model2 overall ~ treatment + cycleno + (1 | randno)      logit
model6 overall ~ treatment + s(cycleno) + (1 | randno)   logit
model4 overall ~ treatment * cycleno + (1 | randno)      logit
model8 overall ~ treatment * s(cycleno) + (1 | randno)   logit


            edf        AIC penlogLik    LR.stat         df Pr(>Chisq)
model1   21.473 -1673.1     849.33                                  
model3   23.481 -1717.3     873.45    48.2406    2.00770  3.397e-11 ***
model5   23.698 -1686.0     856.33   -34.2485    0.21754  1.0000000
model7   27.959 -1719.8     875.38    38.1133    4.26036  1.477e-07 ***
model2  280.048 -3780.9    2031.60  2312.4260  252.08907  < 2.2e-16 ***
model6  282.125 -3783.6    2034.28     5.3778    2.07768  0.0729392 .
model4  282.246 -3788.5    2037.48     6.3931    0.12032  0.0006865 ***
model8  284.282 -3788.5    2037.91     0.8518    2.03688  0.6619574


---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 4 and Model 8 appear to be the best models. As Model 4 is simpler, we prefer it over Model 8. The model parameter estimates are:

```
R> summary(model4)
```

```
Call:
ocm(formula = overall ~ treatment * cycleno + (1 | randno), data = ANZ0001,
    scale = c(0, 100))
```
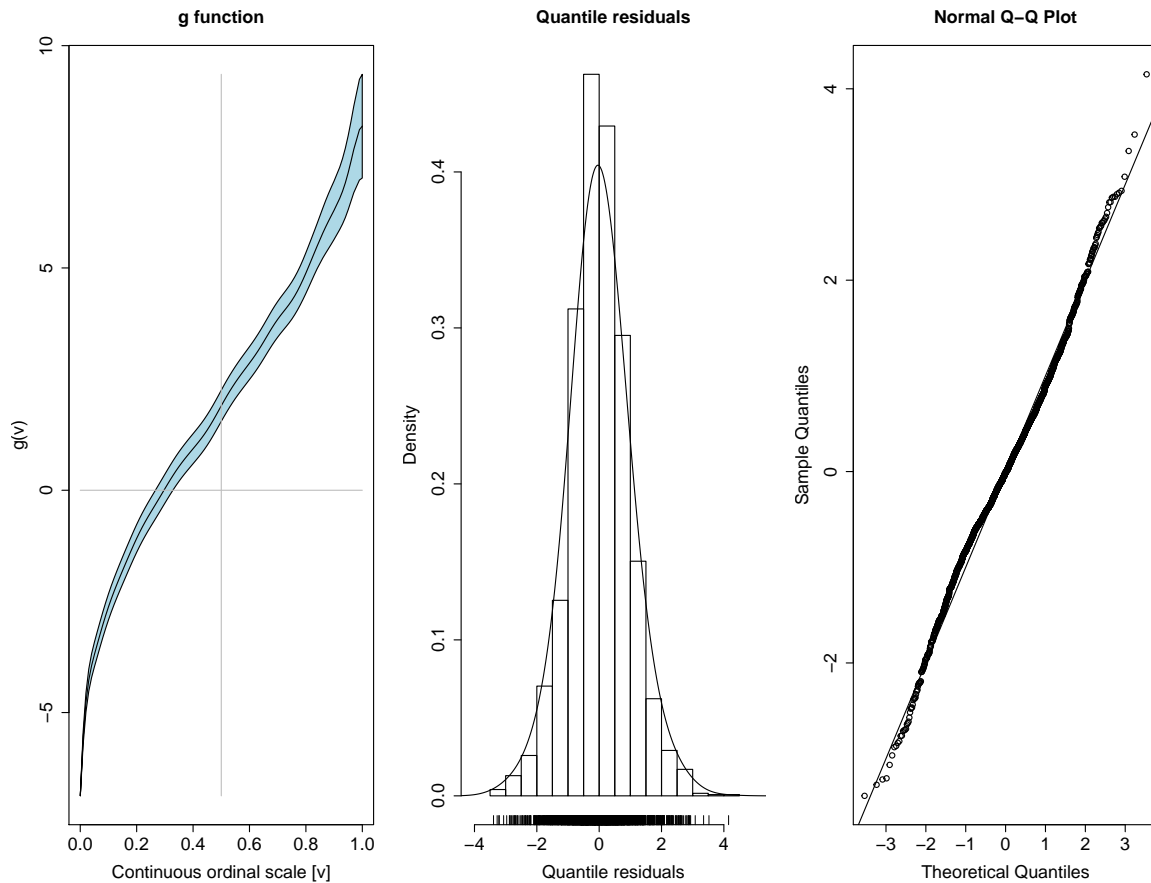
Figure 8: Summary and diagnostic plots for the `model4`, including the estimated g function, the histogram of the quantile residuals and the normal Q-Q plot.

```
Random effects:
            Name Variance Std.Dev.
 Intercept|randno   5.732    2.394

Coefficients:
                     Estimate      StdErr t.value   p.value
treatment2          0.2864264   0.3702171  0.7737   0.43977
treatment3          0.4345589   0.3709857  1.1714   0.24244
cycleno             0.0422235   0.0079961  5.2805 2.576e-07 ***
treatment2:cycleno -0.0121871   0.0110996 -1.0980   0.27315
treatment3:cycleno -0.0785763   0.0269390 -2.9168   0.00382 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can check how well the model fits the data:

```
R> plot(model4)
```

The histogram of the quantile residuals and the Q-Q plot (Figure 8) show that `model4` captures the main sources of variability in the data.

In the following, we show how to use smoothers stratified by a categorical variable, and to do so we study the effect of body-surface area (BSA), treatment and cycle number on the patients' appetite, which, similarly to the overall quality of life, has been measured on a VAS.

Drug dosage in cancer chemotherapy is often based on BSA (Pinkel 1958). BSA can be computed according to a variety of formulae and several studies have analysed the distribution of the different BSA versions in cancer patients. According to Verbraecken, Van de Heyning, De Backer, and Van Gaal (2006) a reasonable value to distinguish female normal weight patients from female overweight patients is 1.68 m^2. In the following, we study the effect of cycle number over appetite for subjects with a BSA lower or higher than 1.68 m^2, according to the model:

$$\text{Model 9:} \quad \sim g(v_{ij}) + t_i + BSA_i + (s(j)|BSA_i) + b_i$$

The R code below creates the categorical variable `BSA_over1.68` and fits the model.

```
R> ANZ0001$bsa_over1.68 = as.factor(ANZ0001$bsa > 1.68)
R> model9  <- ocm(appetite  ~ treatment + bsa_over1.68 +
+    bsa_over1.68:s(cycleno) + (1 | randno), data = ANZ0001,
+    scale = c(0, 100))
R> summary(model9)
R> plot(model9)

Call:
ocm(formula = appetite ~ treatment + bsa_over1.68 + bsa_over1.68:s(cycleno) +
    (1 | randno), data = ANZ0001, scale = c(0, 100))

Random effects:
            Name Variance Std.Dev.
 Intercept|randno    4.711    2.171

Smoother (penalized likelihood smoothing parameter = 1/(2*Std.Dev.)):
                     Name Variance Std.Dev.
 bsa_over1.68FALSE:s(cycleno)    0.002    0.040
  bsa_over1.68TRUE:s(cycleno)    0.000    0.006

Coefficients:
                 Estimate  StdErr t.value p.value
treatment2        0.77094 0.33494  2.3017 0.02209 *
treatment3        0.53409 0.33195  1.6090 0.10876
bsa_over1.68TRUE  0.26458 0.30019  0.8814 0.37887
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can comment that the variable `BSA_over1.68` has not a significant main effect, while the second treatment (CC) is associated with significantly lower values of appetite. Here we can recall Equation 1 and highlight that a positive value of a parameter implies lower values of the outcome (and vice-versa). It is somehow challenging to interpret the estimated values of the parameters, as their direct effect is on the latent and unobserved variable W. In the
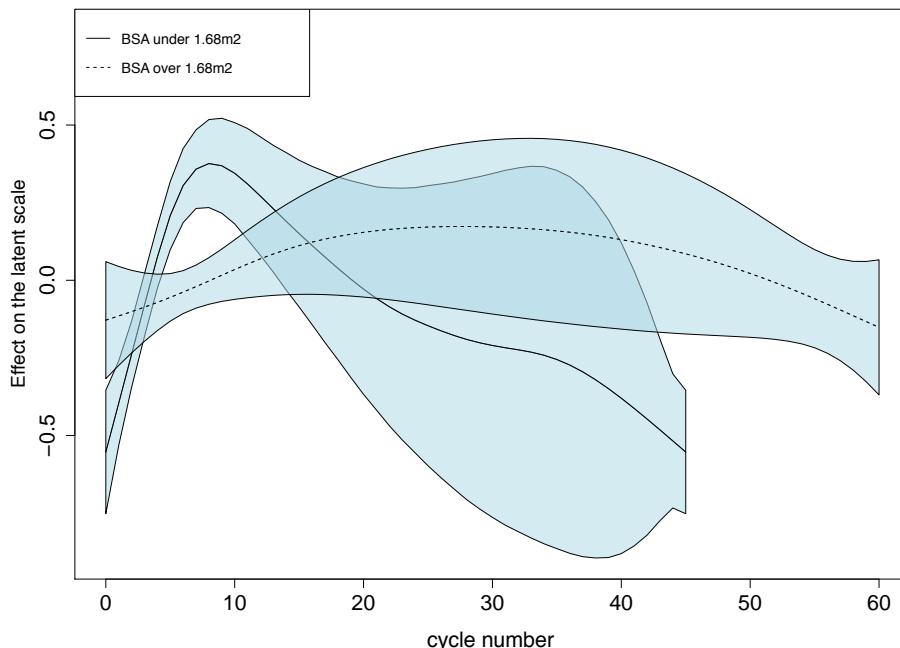
Figure 9: Effect of cycle number on appetite stratified by BSA level, with 95% confidence intervals, as estimated in `model9`. Higher values correspond to lower scores on the continuous ordinal scale.

present case, as we are using the logistic distribution for $\mathsf{F}(\cdot)$, we refer the reader to Section 3.1 and recall that the systematic component $h^\star(X)$ can be interpreted as the log-odds ratio to an observation with $h^\star(X) = 0$, regardless of the value of v. This implies that, keeping the other covariates constant, the odds ratio for subjects who have been treated with the second treatment (CC) to those treated with the first treatment (IC) is $\exp(0.77094) = 2.1618$. The appetite loss associated with the second treatment, relative to the first, is then quantified with a more than two-fold increase in the odds of having less rather than more appetite, independently of the level of appetite. The plot (Figure 9) suggests that patients with low BSA lose their appetite, especially at the beginning of the treatment, with the effect peaking at around the 10th chemotherapy cycle. There is not a significant effect of chemotherapy on the appetite of patients with BSA over 1.68 m^2.

8. Discussion

The package **ordinalCont** implements a regression framework for a response variable that is a recorded perception on a visual analog scale, of an underlying latent variable which is difficult or impossible to observe or measure. The model is an extension of the cumulative logistic ordinal regression model for discrete ordinal responses and is general in the sense of incorporating parametric and smoothing terms, as well as random effects, similarly to what is done in a generalized additive mixed model (GAMM). The recorded perception is mapped to the latent variable using a smooth function, and to avoid overfitting, the model parameters are estimated using constrained optimization of the penalized likelihood. The penalty parameters are automatically selected via maximization of their marginal likelihood.

A commonly used approach for VAS responses is to treat them as ratio variables; while this may in some cases deliver a similar analysis to our approach, this does not recognize the inherent non-ratio and nonlinear nature of the VAS measure. Another approach is to categorize the VAS responses, and analyze them as discrete ordinal responses. While this approach is not incorrect it involves loss of information; our model obviates the need for this aggregation of information.

Computational details

The results in the simulation were obtained on OS X using the reference BLAS libraries shipped with R. The results differ slightly when using a different BLAS, e.g., the vecLib-based BLAS.

References

Atrian MK, Abbaszadeh F, Sarvieh M, Sarafraz N, Jafarabadi MA (2013). "Investigating the Effect of Pressure on Third Liver Point on Primary Dysmenorrhea: a Randomized Controlled Clinical Trial." *Iranian Red Crescent Medical Journal*, **15**(9), 848. `doi:10.5812/ircmj.12719`.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using **lme4**." *Journal of Statistical Software*, **67**(1), 1–48. `doi:10.18637/jss.v067.i01`.

Choi Y, Lee SH, Chung CY, Park MS, Lee KM, Sung KH, Won SH, Lee IH, Choi IH, Cho TJ, Yoo WJ, Lee SY (2014). "Anterior Knee Pain in Patients with Cerebral Palsy." *Clinics in Orthopedic Surgery*, **6**(4), 426–431. `doi:10.4055/cios.2014.6.4.426`.

Chow RT, Heller GZ, Barnsley L (2006). "The Effect of 300 mW, 830 Nm Laser on Chronic Neck Pain: A Double-Blind, Randomized, Placebo-Controlled Study." *Pain*, **124**(1-2), 201–10. `doi:10.1016/j.pain.2006.05.018`.

Christensen RHB (2019). **ordinal**: *Regression Models for Ordinal Data*. R package version 2019.12-10, URL `https://CRAN.R-project.org/package=ordinal`.

Dunn PK, Smyth GK (1996). "Randomized Quantile Residuals." *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. `doi:10.1080/10618600.1996.10474708`.

Hastie T, Tibshirani R (1987). "Non-Parametric Logistic and Proportional Odds Regression." *Journal of the Royal Statistical Society C*, **36**(2), 260–276. `doi:10.2307/2347785`.

Heller GZ, Manuguerra M, Chow R (2016). "How to Analyze the Visual Analogue Scale: Myths, Truths and Clinical Relevance." *Scandinavian Journal of Pain*, **13**, 67–75. `doi:10.1016/j.sjpain.2016.06.012`.

Hothorn T, Möst L, Bühlmann P (2018). "Most Likely Transformations." *Scandinavian Journal of Statistics*, **45**(1), 110–134. doi:10.1111/sjos.12291.

Johnson NL, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions*, volume 2. 2nd edition. John Wiley & Sons.

Kersten P, Küçükdeveci AA, Tennant A (2012). "The Use of the Visual Analogue Scale (VAS) in Rehabilitation Outcomes." *Journal of Rehabilitation Medicine*, **44**(7), 609–610. doi:10.2340/16501977-0999.

Laplace P (1774). "Mémoire Sur La Probabilité Des Causes Par Les évènements." *Mémoires De l'Academie Royale Des Sciences Presentés Par Divers Savan*, **6**, 621–656.

Liu Q, Shepherd BE, Li C, Harrell Jr FE (2017). "Modeling Continuous Response Variables Using Ordinal Regression." *Statistics in Medicine*, **36**(27), 4316–4335. doi:10.1002/sim.7433.

Ma J (2010). "Positively Constrained Multiplicative Iterative Algorithm for Maximum Penalized Likelihood Tomographic Reconstruction." *IEEE Transactions on Nuclear Science*, **57**(1), 181–192. doi:10.1109/tns.2009.2034462.

Manuguerra M, Heller G (2020). **ordinalCont**: *Ordinal Regression Analysis for Continuous Scales*. R package version 2.0.2, URL https://CRAN.R-project.org/package=ordinalCont.

Manuguerra M, Heller G, Ma J (2017). "Semi-Parametric Ordinal Regression Models for Continuous Scales." In M Grzegorczyk, G Ceoldo (eds.), *Proceedings of the 32nd International Workshop on Statistical Modelling*, volume 1, pp. 236–241. University of Groningen.

Manuguerra M, Heller GZ (2010). "Ordinal Regression Models for Continuous Scales." *The International Journal of Biostatistics*, **6**(1). doi:10.2202/1557-4679.1230.

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

Mentula M, Kalso E, Heikinheimo O (2014). "Same-Day and Delayed Reports of Pain Intensity in Second-Trimester Medical Termination of Pregnancy: A Brief Report." *Contraception*, **90**(6), 609–611. doi:10.1016/j.contraception.2014.06.031.

Möst L, Hothorn T (2015). "Conditional Transformation Models for Survivor Function Estimation." *The International Journal of Biostatistics*, **11**(1), 23–50. doi:10.1515/ijb-2014-0006.

Olivan-Blázquez B, Herrera-Mercadal P, Puebla-Guedea M, Pérez-Yus MC, Andrés E, Fayed N, López-Del-Hoyo Y, Magallon R, Roca M, Garcia-Campayo J (2014). "Efficacy of Memantine in the Treatment of Fibromyalgia: A Double-Blind, Randomised, Controlled Trial with 6-Month Follow-up." *Pain*, **155**(12), 2517–2525. doi:10.1016/j.pain.2014.09.004.

Philip BK (1990). "Parametric Statistics for Evaluation of the Visual Analog Scale." *Anesthesia & Analgesia*, **71**(6), 710. doi:10.1213/00000539-199012000-00027.

Pinkel D (1958). "The Use of Body Surface Area as a Criterion of Drug Dosage in Cancer Chemotherapy." *Cancer Research*, **18**(7), 853–856.

Ramsay JO (1988). "Monotone Regression Splines in Action." *Statistical Science*, **3**(4), 425–441. doi:10.1214/ss/1177012761.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richards FJ (1959). "A Flexible Growth Function for Empirical Use." *Journal of Experimental Botany*, **10**(2), 290–301. doi:10.1093/jxb/10.2.290.

Ruppert D, Wand MP, Carroll RJ (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics, 1st edition. Cambridge University Press.

Smithson M, Verkuilen J (2006). "A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables." *Psychological Methods*, **11**(1), 54. doi:10.1037/1082-989X.11.1.54.

Stockler MR, Harvey VJ, Francis PA, Byrne MJ, Ackland SP, Fitzharris B, Van Hazel G, Wilcken NRC, Grimison PS, Nowak AK, Gainford MC, Fong A, Paksec L, Sourjina T, Zannino D, Gebski V, Simes RJ, Forbes JF, Coates AS (2011). "Capecitabine versus Classical Cyclophosphamide, Methotrexate, and Fluorouracil as First-Line Chemotherapy for Advanced Breast Cancer." *Journal of Clinical Oncology*, **29**(34), 4498–4504. doi:10.1200/jco.2010.33.9101.

Svensson E (2000). "Comparison of the Quality of Assessments Using Continuous and Discrete Ordinal Rating Scales." *Biometrical Journal*, **42**(4), 417–434. doi:10.1002/1521-4036(200008)42:4<417::AID-BIMJ417>3.0.CO;2-Z.

Tutz G (2003). "Generalized Semiparametrically Structured Ordinal Models." *Biometrics*, **59**(2), 263–73. doi:10.1111/1541-0420.00033.

Verbraecken J, Van de Heyning P, De Backer W, Van Gaal L (2006). "Body Surface Area in Normal-Weight, Overweight, and Obese Adults. A Comparison Study." *Metabolism*, **55**(4), 515–524. doi:10.1016/j.metabol.2005.11.004.

Wewers ME, Lowe NK (1990). "A Critical Review of Visual Analogue Scales in the Measurement of Clinical Phenomena." *Research in Nursing & Health*, **13**(4), 227–236. doi:10.1002/nur.4770130405.

Yee TW, Wild CJ (1996). "Vector Generalized Additive Models." *Journal of the Royal Statistical Society B*, **58**, 481–493. doi:10.1111/j.2517-6161.1996.tb02095.x.

Affiliation:

Maurizio Manuguerra, Gillian Z. Heller, Jun Ma
Department of Mathematics and Statistics
Faculty of Science and Engineering
Macquarie University
Sydney, NSW 2109, Australia
E-mail: maurizio.manuguerra@mq.edu.au, gillian.heller@mq.edu.au, jun.ma@mq.edu.au
URL: http://maths.mq.edu.au