Statistics and Its Interface Volume 14 (2021) 243–254

Robust regression model for ordinal response

Ao Yuan*, Chongyang Duan, and Ming T. Tan*

Ordinal outcome data with covariates occur frequently in statistical practice including applications from biomedicine to marketing research. Most existing methods for this type of data have relied on subjectively specified models that allow order restriction. There are also some semiparametric ordinal models which are more flexible than parametric ones, with fixed link function, they are still not flexible enough to capture the true link or the relationship between the response and covariates. We propose a broadly applicable robust semiparametric ordinal regression model, in which the relationship between the response and covariates is modelled with a nonparametric monotone increasing link function and parametric regression coefficients. This model is more robust and flexible than existing semiparametric and parametric models for this problem. The semiparametric maximum likelihood estimate is used to estimate the model parameters, and the asymptotic properties of the estimates are derived. Simulation studies show clear advantages of the proposed model over existing parametric models, and a real data analysis illustrates the utility of the proposed method.

Keywords and phrases: Monotone function, Ordinal data, Nonparametric component, Semiparametric maximum likelihood estimate.

1. INTRODUCTION

Ordinal data model with/without covariates is commonly encountered in practice, and has been an area of extensive studies. Existing methods for this type of data have relied on subjectively specified models, and it is known that when the specified model deviates from the true one, the resulting estimates are not reliable. The issue may not be so serious for hypothesis testing and indeed those models have served well applications in, e.g., epidemiological studies where the goal is to explain if certain factors are associated with the outcome. However, such models can be problematic for prediction, as these two goals can be quite different (Shumeli, 2010). For example, in our simulation studies, the true regression coefficients vector is $(0.390, 0.781, -0.488)$, while its estimate under the commonly used logistic model is $(0.682, 1.362, -0.857)$, the estimated regression coefficients are very different but the result of significance test is similar. Thus the predicted values of the ordinal responses are very different. This problem has often escaped our attention. Although there are nonparametric methods that improve the accuracy of modeling by allowing nonlinear covariates, it is difficult to explain the influence of each covariate, thus the semiparametric model is more attractive in practice.

To introduce the problem, we start from the commonly used parametric cumulative logit model (Goodman, 1979; McCullagh, 1980; Anderson, 1984; Laara and Matthews, 1985; Agresti, 1999; Agresti, 2010), for y taking values $\{1, ..., r\}$, $P(y \le r|x) = 1$,

$$P(y \le j|x) = \frac{\exp\{\beta^\intercal x + \alpha_j\}}{1 + \exp\{\beta^\intercal x + \alpha_j\}}, \quad (j = 1, ..., r-1)$$

and so $P(y = j|x) = P(y \le j|x) - P(y \le j-1|x)$, $(1 \le j \le r)$. Alternatively, another common model is, $P(y = r|x) = 1 - \sum_{j=1}^{r-1} P(y = j|x)$, and

$$P(y = j|x) = \frac{\exp\{\beta^\intercal x + \alpha_j\}}{1 + \sum_{k=1}^{r-1} \exp\{\beta^\intercal x + \alpha_k\}}, \quad (j = 1, ..., r-1).$$

The cumulative logit model specifies $\log\left[P(y \le j|x)/P(y > j|x)\right] = \beta^\intercal x + \alpha_j$. There are also various forms of odds ratio models, such as in Fahrmeir and Pritscher (1996). Some authors considered link functions other than exponential form, such as $P(y \le j) = g(\beta^\intercal x + \alpha_j)$ with $g(\cdot)$ being some given distribution function. Chen et al. (1999) considered a class of skewed link models for dichotomous quantal response data. Harville and Mee (1984) investigated a mixed model for ordinal data. Hedeker and Gibbons (1994) and Crouchley (1995) considered a random effects model for ordinal data. Wiliamson et al. (1995) considered a bivariate ordinal data model using odds ratio, Heagerty and Zeger (1996) studied a clustered ordinal data model, Qu and Tan (1998) and Tan et al. (1999) proposed hierarchical Bayesian models for multilevel repeated ordinal data. To achieve more model flexibility and robustness, some authors proposed different forms of semiparametric models for ordinal data. Kauermann and Tutz (2003) and Tutz(2003) proposed model of the form

$$P(y \le j|x, x_1, x_2) = F\big(\beta_j + \beta^\intercal x_1 + \gamma(x) + x_2^\intercal \gamma_z(x)\big)$$

or

$$P(y = j|y \ge j, x, x_1, x_2) = F\big(\beta_j + \beta^\intercal x_1 + \gamma(x) + x_2^\intercal \gamma_z(x)\big)$$

where $F(\cdot)$ is a known link function, and $\gamma(\cdot)$ and $\gamma_z(\cdot)$ are unknown functions. They use kernel smoothing and pro-

*Corresponding authors.

file likelihood methods to estimate the nonparametric components. Das et al. (2016) considered the logit link for a partial linear regression model. Adeyemi et al. (2016) used this model in the analysis of child birth weight. The model proposed by Kazembe (2009) has a similar form. Sutradhar (2017) proposed a semiparametric model for longitudinal ordinal data. Xia and Gou (2016) studied a Bayesian semiparametric model for mixed continuous and ordinal outcomes. Kottas et al. (2005) considered a nonparametric Bayesian model for multivariate ordinal data. For a review of methods in this field, see Kauermann and Tutz (2003).

Almost all existing models assume that the link function is known or known up to a class of distributions. To have greater model robustness and flexibility, we propose a semiparametric model different from the existing ones. For an ordinal response variable y taking values in $\{1, ..., r\}$ and covariates x, we specify $P(y \leq r|x, \lambda, F) = 1$, and

$$P(y \leq j|x, \lambda, F) = F(\beta^\intercal x + \lambda_j), \quad (j = 1, ..., r-1),$$

where β is a vector of regression coefficients, λ_j is the effect of the j-th cumulative ordinal response, and $F(\cdot)$ is an unknown monotone increasing function, in particular a cumulative distribution function (CDF), as opposed to any fixed link functions. This model is more flexible than the ones above mentioned. The model parameters are estimated by the semiparametric maximum likelihood estimate.

The proposed approach utilizes nonparametric/semiparametric models with shape constraint on the nonparametric component. The advantage of this class of models is that the estimated nonparametric component is well behaved and its asymptotic distribution can be studied. It has applications in many real problems. For earlier works on this topic, see Grenander (1956), Prakasa Rao (1969, 1970), Kiefer and Wolfowitz (1976), Lo (1986), Robertson et al. (1988), Bickel and Fan (1996), Birgé (1997), among others. For recent applications using this method, see Huang (2002), Stout (2008), Qin et al. (2014), Turnball and Ghosh (2016), Yuan, Yin and Tan (2019), Yuan et al. (2019), among others. The semiparametric model for binary data is studied in Cosslett (1983), Klein and Spady (1993). The purpose of this article is to develop a shape constrained semiparametric model for ordinal data, prove its asymptotic properties and assess its finite sample performance.

The rest of the paper is organized as follows. In Section 2, we describe the proposed model and its parameter estimation. Section studies the asymptotic behavior of the parameter estimates; Section 4 conducts simulation studies to evaluate performance of the proposed method, and compare it with the commonly used parametric models; The proposed method is used to analyze a real ordinal data set in Section 5, and we conclude with remarks in Section 6. All relevant proofs are given in the Appendix.

2. THE PROPOSED METHOD

Let the observed ordinal data be $D_n = \{(y_i, x_i) : i = 1, ..., n\}$ from n independent individuals, where y_i is the ordinal response of the i-th subject, it can take values $1, ..., r$ for some small integer r (often $3 \leq r \leq 10$); and $x_i \in R^d$ is the corresponding covariate vector.

2.1 The model specification

Let $\beta = (\beta_1, ..., \beta_d)^\intercal$ be the regression coefficients of the covariate x_i's. We propose the following semiparametric model:

$$P(y \leq j|x, \lambda, F) = F(\beta^\intercal x + \lambda_j), \quad (j = 1, ..., r), \theta \in \Theta, \ F \in \mathcal{F} \tag{1}$$

with the constraint $\|\beta\| = 1$ for model identifiability, where $0 = \lambda_1 < ... < \lambda_{r-1} < \lambda_r = +\infty$ are parameters correspond to the ordinal responses, $\lambda = (\lambda_2, ..., \lambda_{r-1})^\intercal$, $\theta = (\beta^\intercal, \lambda)^\intercal$, Θ is a subset of R^d ($d = dim(\theta)$), $\mathcal{F}$ is the collection of all distribution functions on R. The specified model is also parsimonious and constrains $F(\cdot)$ to be monotone. Our goal is to derive semiparametric maximum likelihood estimates of the model parameters (θ, F).

The likelihood for the observed data is

$$\prod_{i=1}^{n} F^{I(Y_i=1)}\big(\beta^\intercal x_i + \lambda_1\big) \prod_{j=2}^{r} \Big(F\big(\beta^\intercal x_i + \lambda_j\big) - F\big(\beta^\intercal x_i + \lambda_{j-1}\big)\Big)^{I(Y_i=j)}.$$

However, computation of the nonparametric maximum likelihood estimate (NPMLE) of F from the above likelihood is difficult, so instead we use the following pseudo likelihood

$$\prod_{i=1}^{n}\prod_{j=1}^{r} F^{I(Y_i \leq j)}\big(\beta^\intercal x_i + \lambda_j\big)\Big(1 - F\big(\beta^\intercal x_i + \lambda_{j-1}\big)\Big)^{1-I(Y_i \leq j)}.$$

Letting $t_{ij} = I(Y_i \leq j)$ $(j = 1, ..., r)$, the corresponding pseudo log-likelihood is

$$\ell_p(\theta, F|D_n) = \sum_{i=1}^{n}\sum_{j=1}^{r} \Big(t_{ij} \log F(\beta^\intercal x_i + \lambda_j) + (1 - t_{ij}) \log\big(1 - F(\beta^\intercal x_i + \lambda_{j-1})\big)\Big). \tag{2}$$

Then we can estimate the parameters by the semiparametric pseudo MLE $(\hat{\theta}, \hat{F})$,

$$(\hat{\theta}, \hat{F}) = \arg \max_{(\theta, F) \in (\Theta, \mathcal{F})} \ell_p(\theta, F|D_n). \tag{3}$$

The pseudo MLE (also known as quasi MLE) may not be exactly the MLE under the 'true' likelihood, but it still enjoys similar asymptotic properties of the latter, and is broadly studied and used in statistical inferences, such as in Wedderburn (1974), McCullagh and Nelder (1989), Cox and Reid (2004), and Amini et al. (2013).

2.2 Computation of parameter estimation

The maximization in formula 3 is not trivial. Here we describe an iterative maximization procedure for its computation:

1. For a given starting value $\theta^{(0)}$ of θ, find $F^{(1)}(\cdot) \in \mathcal{F}$ as the maxima of $\ell_p(\theta^{(0)}, F|D_n)$;
2. Then fix $F^{(1)}$, find $\theta^{(1)} \in \Theta$ as the maxima of $\ell_p(\theta, F^{(1)}|D_n)$;
3. Repeat the two steps until the sequence $\{(\theta^{(r)}, F^{(r)})\}$ converges.

It is easy to see that the likelihood as function of the sequence $\{(\theta^{(r)}, F^{(r)})\}$ increases as the iteration r goes to the next, and will converge to at least some local maxima of $\ell_p(\theta, F|D_n)$. In other words, for all integer r,

$$\begin{aligned}\ell_p(\theta^{(r+1)}, F^{(r+1)}|D_n) &\geq \ell_p(\theta^{(r)}, F^{(r+1)}|D_n)\\ &\geq \ell_p(\theta^{(r)}, F^{(r)}|D_n).\end{aligned}$$

Thus, the sequence $\{(\theta^{(r)}, F^{(r)})\}$ will converge to some stationary point of ℓ_p, which may or may not be the global maxima. To locate the global maxima, multiple starting values are needed, and if they resulted in several different stationary points, we choose the one which gives the largest pseudo likelihood value. This issue is similar to that with the EM algorithm, and is common in semiparametric estimation.

However, the maximization over F is non-standard. For brevity, we just denote $F^{(r)}$ by $\hat{F}$, fix the value of θ, and let $z_{ij} = \beta^\intercal x_i + \lambda_j$. We use the technique of isotonic regression for the above maximization. To do this, we need to re-write the pseudo likelihood into a sum of squares. Recall the maximization problem for the current status model, such as in Example 3.2.15 of van der Vaart and Wellner (1996, p. 298), the maximization in formula 3 is written as the following minimization

$$\hat{F} = \arg\min_{F\in\mathcal{F}} \sum_{i=1}^{n}\sum_{j=1}^{r}\big(F(z_{ij}) - t_{ij}\big)^2. \tag{4}$$

The above minimization subject to the monotone constraint on F then becomes the standard isotonic regression problem, and can be computed by the pool adjacent violators algorithm (PAVA), see for example, Best and Chakravarti (1990) using the R-package *isoreg* or *ufit*. Patric et al. (2009) reviewed the history of this algorithm and computation aspects.

However, like any NPMLE, $\hat{F}$ is a dis-continuous step function, with masses concentrating on some of the observation points. It may be desirable to have a continuous version of $\hat{F}$ as below. Given a kernel $K(\cdot)$ (a density function), let $\mathbb{K}(s) = \int^s K(u)du$ be the CDF of $K(\cdot)$, then the following modified distribution function estimator is continuous

$$\tilde{F}(s) = \int \mathbb{K}\big(\frac{s-u}{h}\big)d\hat{F}(u),$$

typically with bandwidth $h = O(n^{-1/5})$. Here $\tilde{F}(\cdot)$ is continuous, monotone increasing, and enjoys nice asymptotic properties as $\hat{F}$. Letting $\{s_1, ..., s_m\}$ $(m \leq n)$ be the atoms of $\hat{F}$, then $\tilde{F}$ is re-written as

$$\tilde{F}(s) = \sum_{j=1}^{m}\mathbb{K}\big(\frac{s-s_j}{h}\big)\big(\hat{F}(s_j) - \hat{F}(s_j - 0)\big).$$

2.3 Hypothesis test

In practice, for example in clinical trials, we want to compare responses between patients received a new treatment and a control therapy, and the interest is to test the treatment difference in response /outcome. Denote the regression coefficients of treatment vs control as β. We want to test the null hypothesis $H_0 : \beta = 0$ vs the alternative $H_1 : \beta \neq 0$. Let $\hat{\beta}$ be the estimated values of β based on data. For ease of exposition, we assume the two groups have the same sample size n. We use the Wald type statistic for this problem as $T_n = \hat{\beta}^2 Var^{-1}(\hat{\beta})$. The variance matrix $Var(\hat{\beta})$ is computed by the bootstrap, and for a given significance level α (typically $= 0.05$), the critical value is computed via simulation under H_0.

3. ASYMPTOTIC PROPERTIES OF PARAMETER ESTIMATION

To study the asymptotic behavior of the estimators, we use the following notations. Let $\mathcal{B}$ be the range of β, D be the support of $\beta^\intercal x + \lambda_j$'s, and (θ_0, F_0) be the Kullback-Leibler projection of the true conditional distribution onto the pseudo parameter space $(\Theta, \mathcal{F})$ in the sense

$$(\theta_0, F_0) = \arg\min_{(\theta,F)\in(\Theta,\mathcal{F})} E\big\{\log\frac{P(y|x)}{F(\beta^\intercal x + \lambda_y)}\big\},$$

where $P(y \leq j|x)$ is the true conditional mass function of y given x, and E is expectation under the true joint distribution of (y, x). Under some mild conditions such that (θ_0, F_0) exists (see, Pfanzagl, 1990; Patilea, 2001), and since the model $F(\beta^\intercal x + \lambda_j)$ is identifiable, it is unique. If $P(y|x) \in \{F(\beta^\intercal x + \lambda_y) : (\theta, F) \in (\Theta, \mathcal{F})\}$, (θ_0, F_0) will be the true parameters generating the observed data. This is analogous to the true parameters of a semiparametric model which is also the Kullback-Leibler projection of the true model. We first list four relatively simple conditions below.

C1. $\theta_0 \in \Theta$, which is compact.
C2. The support of X is compact.
C3. $\max_{1\leq j\leq r}\sup_{(\theta,F)\in(\Theta,\mathcal{F})} E\big[F(\beta^\intercal X + \lambda_j)\big(1 - F(\beta^\intercal X + \lambda_j)\big)\big]^{-2} < \infty$.
C4. $\forall F \in \mathcal{F}$, $F(\cdot)$ is differentiable and uniformly bounded, and $F_0 \in \mathcal{F}$

Then we have the following theorem for the properties of the semiparametric pseudo MLE $(\hat{\theta}, \hat{F})$ (all relevant proofs are given in the Appendix).

Theorem 3.1. *Assuming (C1)–(C4), then as* $n \to \infty$,

$$\hat{\theta} \overset{a.s.}{\to} \theta_0, \qquad \sup_{s\in D} |\hat{F}(s) - F_0(s)| \overset{a.s.}{\to} 0.$$

Proof. Provided in Appendix A. □

The asymptotic distribution and the exact convergence rate of $\hat{\theta}$ are open questions. As pointed out in Huang and Wellner (1997, Section 3.2.2), Murphy, van der Vaart and Wellner (1999; Sections 2–3) and Groeneboom and Hendrickx (2018), in their models, the $\sqrt{n}$-consistency of the MLE $\hat{\theta}$ is still unknown. The reason is that $\hat{\theta}$ and $\hat{F}$ are bundled together, and that $\hat{F}(\cdot)$ is non-smooth. So below we only give the convergence rate of $\|\hat{\theta} - \theta_0\| + \|\hat{F} - F_0\|$, with $\|\hat{\theta} - \theta_0\|$ being the Euclidean distance between $\hat{\theta}$ and θ_0, and $\|\hat{F} - F_0\| = \sup_{s\in D} |\hat{F}(s) - F_0(s)|$.

Theorem 3.2. *Under conditions (C1)–(C4),*

$$\|\hat{\theta} - \theta_0\| + \|\hat{F} - F_0\| = O_p(n^{-1/3}).$$

Proof. Provided in Appendix A. □

Next we give the asymptotic distribution of $\hat{F}$. Let $\dot{F}_0(\cdot)$ be the derivative of $F_0(\cdot)$, $\mathbb{B}(\cdot)$ be the two-sided Brownian motion originating from zero: a mean zero Gaussian process on R with $\mathbb{B}(0) = 0$, and $E\big(\mathbb{B}(s) - \mathbb{B}(h)\big)^2 = |s - h|$ for all $s, h \in R$. Although the $\sqrt{n}$ convergence of $\hat{\theta}$ is still open, numerical studies in ours and related papers suggest that its convergence rate is faster than $n^{1/3}$, so it is reasonable to assume $\|\hat{\theta} - \theta_0\| = o_p(n^{-1/3})$. Let $\overset{D}{\to}$ denote convergence in distribution. With two additional conditions

C5. $S := \beta_0^\intercal X$ has a density $g(\cdot)$ with $\bar{g}(s) := (r-1)^{-1}\sum_{j=1}^{r-1} g(s - \lambda_{0,j}) > 0$ for some s in the support of S.

C6. $\|\hat{\theta} - \theta_0\| = o_p(n^{-1/3})$.

we obtain

Theorem 3.3. *Assuming (C1)–(C6) and denoting* $\dot{F}_0(s) = dF_0(s)/ds$, *we have*

$$n^{1/3}(\hat{F}(s) - F_0(s)) \overset{D}{\to} \Big(\frac{4\eta^2(s)\dot{F}_0(s)}{\bar{g}^2(s)}\Big)^{1/3} \arg\min_{h\in R}\{\mathbb{B}(h) + h^2\},$$

where $\eta^2(s) = (r-1)^{-2}\sum_{j=1}^{r-1} E\big[\big(I(Y \le j) - F_0(s)\big)^2 \big| S = s - \lambda_{0,j}\big] g(s - \lambda_{0,j})$

Proof. Provided in Appendix A. □

Table 1. Parameter estimates under three models (simulated data)

Data	$\boldsymbol{\theta}$	$\boldsymbol{\beta}$ *for covariates*	$\boldsymbol{\beta}$ *for group*	$\boldsymbol{\lambda}$
F_0 = Normal(0,1)	$\boldsymbol{\theta}_0$	(0.390,0.781,-0.488)	(0.000)	(0.000,0.500)
$m = 3$	Robust model	(0.387,0.771,-0.485)	(0.009)	(0.000,0.501)
$r = 3$	[sd]	[0.038,0.033,0.030]	[0.132]	[,0.083]
	Logistic model	(0.686,1.372,-0.855)	(0.014)	(-0.006,0.867)
	[sd]	[0.079,0.118,0.077]	[0.227]	[0.166,0.173]
	Probit Model	(0.398,0.795,-0.496)	(0.008)	(0.002,0.503)
	[sd]	[0.044,0.065,0.043]	[0.131]	[0.096,0.099]
F_0 = Beta(5,1)	$\boldsymbol{\theta}_0$	(0.379,0.758,-0.474)	(0.237)	(0.000,1.000,1.500,2.500)
$m = 3$	Robust model	(0.378,0.756,-0.473)	(0.239)	(0.000,0.952,1.452,2.458)
$r = 3$	[sd]	[0.021,0.019,0.018]	[0.067]	[,0.078,0.089,0.110]
	Logistic model	(0.820,1.642,-0.616)	(0.516)	(-1.841,0.010,1.018,3.182)
	[sd]	[0.069,0.111,0.059]	[0.184]	[0.199,0.161,0.148,0.197]
	Probit Model	(0.466,0.933,-0.350)	(0.294)	(-0.985,0.072,0.652,1.879)
	[sd]	[0.039,0.062,0.034]	[0.106]	[0.110,0.090,0.084,0.109]
F_0 = Exp(1)	$\boldsymbol{\theta}_0$	(0.270,0.338,0.405,-0.608,-0.473)	(0.236)	(0.000,1.000,1.800,2.800)
$m = 5$	Robust model	(0.268,0.336,0.405,-0.606,-0.471)	(0.235)	(0.000,1.009,1.799,2.724)
$r = 5$	[sd]	[0.021,0.026,0.020,0.022,0.032]	[0.069]	[,0.073,0.090,0.120]
	Logistic model	(0.527,0.662,0.793,-1.189,-0.925)	(0.456)	(-1.748,0.346,1.898,3.642)
	[sd]	[0.064,0.075,0.067,0.087,0.107]	[0.188]	[0.159,0.164,0.214,0.295]
	Probit Model	(0.286,0.359,0.429,-0.645,-0.501)	(0.247)	(-1.022,0.126,0.970,1.899)
	[sd]	[0.038,0.044,0.039,0.051,0.062]	[0.112]	[0.091,0.093,0.118,0.161]
F_0 = Beta(0.6,0.6)	$\boldsymbol{\theta}_0$	(0.278,0.348,0.417,-0.626,-0.487)	(0.000)	(0.000,1.500)
$m = 5$	Robust model	(0.267,0.333,0.400,-0.602,-0.461)	(-0.003)	(0.000,1.422)
$r = 3$	[sd]	[0.065,0.085,0.065,0.070,0.107]	[0.228]	[,0.186]
	Logistic model	(0.184,0.233,0.277,-0.421,-0.326)	(0.001)	(-0.022,1.011)
	[sd]	[0.056,0.069,0.050,0.055,0.103]	[0.188]	[0.138,0.144]
	Probit Model	(0.115,0.146,0.173,-0.263,-0.203)	(0.000)	(0.004,0.622)
	[sd]	[0.033,0.041,0.029,0.033,0.060]	[0.111]	[0.082,0.085]

F_0 =True distribution, m =No. of Co-variables, r =No. of Orders

For the smoothed version $\tilde{F}(\cdot)$ of $\hat{F}(\cdot)$ given at the end of Section 2.2, using the theory of Groeneboom and Jongbloed (2014, Theorem 11.4, p. 332–335) and under suitable conditions, we can show that $n^{2/5}\big(\tilde{F}(s) - F_0(s)\big) \xrightarrow{D} N(\mu, \sigma^2)$, for some finite (μ, σ^2).

4. SIMULATION STUDY

To evaluate model performance, we simulated four data sets with different combinations of number of orders (r), numbers of covariates (m), different parameters (θ_0) and different link functions (F). In each set, we simulate $n = 500$ i.i.d. data (y_i, x_i) $(i = 1, ..., n)$, with number of orders $r = 3$ or 5, and with covariates $x_i = (x_{i1}, ..., x_{im})$, $m = 3$ or 5 and with one group variable (treatment vs. control). We first generate the covariates, sample the x_i's for the covariates from the 3 (or 5)-dimensional normal distribution $N(\mu, \Sigma)$ and that for the group variable from the binomial distribution. Without loss of generality, we set the mean vector $\mu = \mathbf{0}$ and some given covariance matrix Γ.

Then we estimate the parameters θ from model 3, for some different choices of θ_0, compare the corresponding estimates with the classical models with the *logistic* link and *probit* link for this problem. For the starting value of our robust model, the estimation from several classical parametric models with best AIC is used. The results are given in Table 1 below, with estimated standard deviation (sd) in square bracket. We also computed the test statistic described in Section 2.3, and compared the results from the logistic and probit models, with type I errors and powers given in Table 2 below. The estimated standard deviations, type I errors and powers are all computed by 1000 repetitions.

The estimated curve $\hat{F}(\cdot)$ vs the true ones are shown in Figure 1. We see that the nonparametric estimate $\hat{F}$ and its smooth version $\tilde{F}(\cdot)$ fit the true link function very well, but those from the *logistic* and *probit* links deviate from true one significantly, except the case the *probit* link is the true one in the simulation (the first panel of Figure 1).

From Table 1, we see that estimates from the proposed method have better overall performance than those from the classical *logistic* or *probit* link model when the latter departs significantly from the true link. In particular, estimates under the proposed model are much more accurate than those under the *logistic* or *probit* link model, and with smaller standard deviations. For hypothesis testing, statistical tests from the three models have comparable type I error (for the first data set in Table 2, the probit link is the true one in the simulation). However, the proposed model has significantly higher power than the other two. Generally, as the sample size increases, the estimator behaves better with smaller bias and estimated standard deviation. In fact, as the sample size increased to 1000 in our simulation, smaller biases and estimated standard deviations are observed, even with more parameters to estimate (results not shown here due to space limitation).

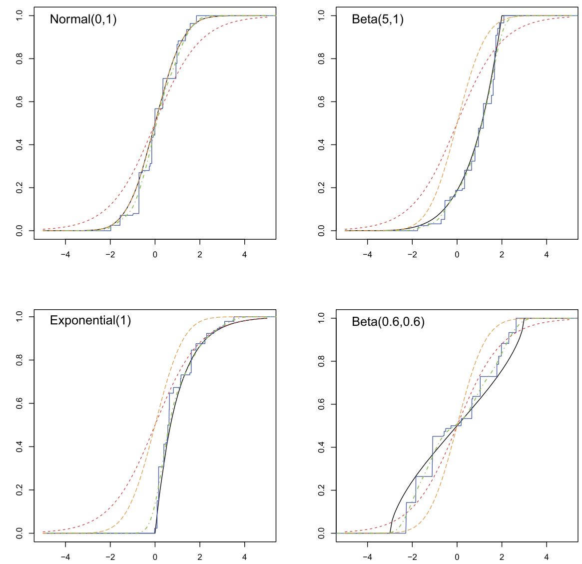

Figure 1. $\hat{F}, \tilde{F}$ vs F_0. solid line – true link function F_0; dotted line – estimated logit link; dash line – estimated probit link; step line – $\hat{F}$; dash dot line – $\tilde{F}$.

Table 2. Parameter estimates test under three models (simulated data)

Data	Model	Type I error or Power
F_0=Normal(0,1)	Logistic model	*Type I error* = 0.066
$m = 3$	Probit model	*Type I error* = 0.057
$r = 3$	Robust model	*Type I error* = 0.065
F_0=Beta(5,1)	Logistic model	*Power* = 0.810
$m = 3$	Probit model	*Power* = 0.804
$r = 3$	Robust model	*Power* = 0.950
F_0=Exp(1)	Logistic model	*Power* = 0.695
$m = 5$	Probit model	*Power* = 0.633
$r = 5$	Robust model	*Power* = 0.827
F_0=Beta(0.6,0.6)	Logistic model	*Type I error* = 0.042
$m = 5$	Probit model	*Type I error* = 0.036
$r = 3$	Robust model	*Type I error* = 0.043

F_0 =True distribution, m =No. of Co-variables, r =No. of Orders

5. APPLICATION TO REAL TRIAL

Now we use the proposed model to analyze the following real data set from the HORIZON III study (https://www.projectdatasphere.org/projectdatasphere/html/content/78). It is a randomized, double-blind, multicenter adaptive phase II/III study to compare the efficacy of Cediranib plus FOLFOX6 versus Bevacizumab plus FOLFOX6 in patients with previously untreated metastatic colorectal cancer. In phrase II, patients were randomly assigned one of the two treatments. An independent end-of-phase II analysis concluded that mFOLFOX6/cediranib 20 mg met predefined criteria for continuation; subsequent patients received mFOLFOX6/Cediranib 20 mg or mFOLFOX6/Bevacizumab (randomly assigned 1:1). In all, 1,422 patients received mFOLFOX6/Cediranib 20 mg (n = 709)

Table 3. Parameter estimates from three models for the real data

Variables	HORIZON III clinical trial					
	$\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ (sd)					
	Robust model	P values	Logistic model	P values	Probit model	P values
Age, 5 years	-0.030 (0.038)	0.465	-0.026 (0.039)	0.505	-0.016 (0.023)	0.490
Sex, male vs. female	-0.345 (0.164)	0.029	-0.382 (0.171)	0.025	-0.219 (0.101)	0.031
WHO performance status, normal activity vs. others	0.655 (0.128)	0.000	0.588 (0.168)	0.000	0.360 (0.100)	0.000
Adjuvant therapy, yes vs. no	0.057 (0.239)	0.781	-0.187 (0.257)	0.469	-0.077 (0.157)	0.624
Tumor grading, poorly differentiated vs. others	-0.481 (0.217)	0.022	-0.271 (0.225)	0.229	-0.191 (0.131)	0.145
VEGF, $\geq$ 98 vs. <98 pg/mL	-0.376 (0.164)	0.020	-0.325 (0.168)	0.053	-0.198 (0.100)	0.048
Albumin, 5 g/L	0.016 (0.075)	0.831	-0.044 (0.071)	0.530	-0.017 (0.041)	0.672
LDH, log transformed	-0.137 (0.131)	0.295	-0.056 (0.122)	0.645	-0.047 (0.071)	0.502
ALP, log transformed	-0.238 (0.136)	0.051	-0.313 (0.129)	0.015	-0.175 (0.077)	0.023
λ_1	–		0.098 (0.187)		0.053 (0.112)	
λ_2	2.336 (0.399)		2.347 (0.219)		1.368 (0.123)	

VEGF, vascular endothelial growth factor; LDH, lactate dehydrogenase; ALP, alkaline phosphatase.

or mFOLFOX6/Bevacizumab (n = 713). The primary objective was to compare progression-free survival (PFS) and the primary analysis revealed no significant difference between the two arms in PFS (hazard ratio[HR], 1.10; 95% CI, 0.97 to 1.25; P = 0.119). In the present example, only mFOLFOX6/Bevacizumab group data is available. The outcome of best overall response (3 for progressive disease (PD), 2 for stable disease [SD], 1 for complete response (CR) + partial response (PR)) is analyzed and the available variables (WHO performance status, Albumin, Alkaline Phosphatase, Tumor Grade, Vascular Endothelial Growth Factor, Lactate Dehydrogenase, Age, Male, Adjuvant therapy) that considered by the authors to do subgroup analysis are included in the model. After excluding patients with missing data which are assumed to be missing completely at random, 581 patients are included in this analysis. The estimated $\hat{F}$ by using our proposed robust model is different from the *logistic distribution* and *normal distribution* (Figure 2). The parameter estimates are shown in following Table 3. The results of our robust model, the *logistic* link model and the *probit* link model are similar, but tumor grading and vascular endothelial growth factor (VEGF) are significant by using the robust model while not by using the *logistic* link model. Clinically, since tumor grading is a well known risk factor and VEGF is up-regulated in many tumors and its contribution to tumor angiogenesis is well defined, so the results by using the robust model are consistent with existing clinical evidence. Statistically, since the estimated curve and the logistic and normal curves do not match very well, the results of the robust model are more reliable.

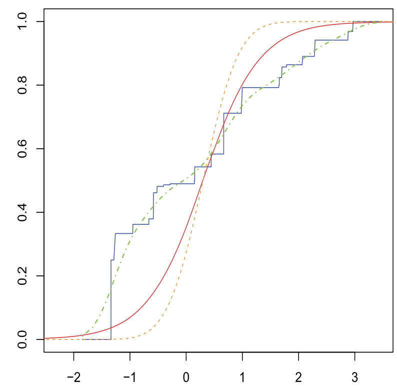

Figure 2. $\hat{F}$ (blue step line), $\tilde{F}$ (green dash dot line) vs logistic (red solid line) and normal (orange dotted line) distributions in real data.

6. CONCLUDING REMARKS

We have proposed a robust semiparametric ordinal model with a nonparametric link function for analyzing ordinal data. The regression model retains the parametric component for explicit assessment of certain covariates while the nonparametric components provide model flexibility and robustness. We derived the semiparametric maximum likelihood estimate for model parameter estimation, using the isotonic regression method, and its asymptotic properties. Simulation studies show that the proposed model outperforms the commonly used logistic model for this problem when the latter deviates from the true link. The method is then used to analyze a real ordinal dataset. As a semiparametric model with both parametric and nonparametric parts, its implementation is more complicated than that of the commonly used parametric logistic model. For example, using the existing R-package *isoreg*, the computation time of the proposed model for the real data set is about 0.21036 seconds in a typical laptop. In comparison the computation time for the MLE under the classical logistic model (R *polr* function) takes about 0.07547 seconds. So the added complexity is trivial and not a hindrance for applying the proposed method, and worthwhile given its apparent advantages. In conclusion, the proposed new class of models provide a flexible, robust and useful method for analyzing ordinal data. It is easy to use and added computational complexity is well manageable.

APPENDIX A. PROOFS AND ASYMPTOTICS

A.1 Proof of Theorem 1

Let $g(\cdot)$ be the density of X,

$$F(j|x,\beta) = F(\beta^{\intercal}x + \lambda_j), (\lambda_0, \lambda_r) = (0, \infty).$$

Denote

$$f(y|x,\theta) = \prod_{j=1}^{r} F^{I(y\leq j)}(\beta^{\intercal}x+\lambda_j)\big(1-F(\beta^{\intercal}x+\lambda_j)\big)^{1-I(y\leq j)},$$

$$p(y,x|\theta,F) = f(y|x,\theta)g(x).$$

The true joint mass/density of (Y,X) is $p(y,x) = f(y|x)g(x)$. Let

$$q(y,x|\theta,F) = \log\frac{p(y,x)}{p(y,x|\theta,F)}$$

be the pseudo log-likelihood ratio. Let P be the probability measure of $p(y,x)$,

$$Pq(Y,X|\theta,F) = \int q(y,x|\theta,F)dP(y,x),$$

the true mean of q; and let

$$P_n q(Y,X|\theta,F) = (1/n)\sum_{i=1}^{n} q(y_i,x_i|\theta,F),$$

the empirical mean of q based on the data $\{(y_i,x_i) : i = 1,...,n\}$.

Note that $Pq(Y,X|\theta,F)$ is the negative Kullback-Leibler divergence of $p(y,x|\theta,F)$ from $p(y,x)$, and as a function of (θ,F), it is always non-positive, attaining its maximum value at $(\theta,F) = (\theta_0,F_0)$. Since $(\hat{\theta},\hat{F})$ is the pseudo MLE of (θ_0,F_0),

$$(\theta_0,F_0) = \arg\max_{(\theta,F)\in(\Theta,\mathcal{F})} Pq(Y,X|\theta,F),$$

and

$$(\hat{\theta},\hat{F}) = \arg\max_{(\theta,F)\in(\Theta,\mathcal{F})} P_n q(Y,X|\theta,F).$$

Denote

$$d(\theta,F;\theta_0,F_0) = \|\theta-\theta_0\| + \sup_{s\in D}|F(s)-F_0(s)|.$$

By our model specification, $Pq(Y,X|\theta,F)$ is continuous with respect to (θ,F); also, the model is identifiable, so (θ_0,F_0) is the unique maximizer of $Pq(Y,X|\theta,F)$. Thus for all $\eta > 0$,

$$\sup_{(\theta,F)\in(\Theta,\mathcal{F}):d(\theta,F;\theta_0,F_0)>\eta} Pq(Y,X|\theta,F) < Pq(Y,X|\theta_0,F_0)$$

and

$$P_n q(Y,X|\hat{\theta},\hat{F}) \geq P_n q(Y,X|\theta_0,F_0),$$

so if we show that $\mathcal{Q} = \{q(y,x|\theta,F) : (\theta,F) \in (\Theta,\mathcal{F})\}$ is a P-Glivenko-Cantelli class, then by Theorem 5.8 in van der Vaart (2002, p. 386),

$$d(\hat{\theta},\hat{F};\theta_0,F_0) \overset{a.s.}{\to} 0.$$

This gives the desired result.

Now we show that $\mathcal{Q}$ is P-Glivenko-Cantelli. For any function g, let $\|g\|_{L_1(P)} = \int|g(y,x)|dP(y,x)$, and $N_{[\,]}(\epsilon,\mathcal{Q},L_1(P))$ be the minimum number of ϵ-brackets needed to cover $\mathcal{Q}$ under norm $\|\cdot\|_{L_1(P)}$. We first show that $N_{[\,]}(\epsilon,\mathcal{Q},L_1(P))$ is finite $\forall\epsilon > 0$.

By our specification of $p(y,x|\theta,F)$ and with (C1) and (C3), it can be checked that $q(y,x|\theta,F)$ is boundedly differentiable with respect to θ, and boundedly Gáteaux differentiable with respect to F. In fact, let

$$\dot{\ell}_{p,\theta}(\theta,F|y,x) = \partial\ell_p(\theta,F|y,x)/\partial\theta$$

be the score for θ, and

$$\dot{\ell}_{p,F}(\theta,F|y,x)[h] = \partial\ell_p(\theta,F+\lambda h|y,x)/\partial\lambda\big|_{\lambda=0}$$

be the Hadamard derivative of $\ell_p(\theta,F|y,x)$ with respect to F in the direction h. Then for some (θ,F) lies between (θ_1,F_1) and (θ_2,F_2),

$$\begin{aligned} &q(y,x|\theta_2,F_2) - q(y,x|\theta_1,F_1)\\ &= \ell_p(y,x|\theta_2,F_2) - \ell_p(y,x|\theta_1,F_1)\\ &= \dot{\ell}_{p,\theta}(\theta,F|y,x)(\theta_2-\theta_1) + \dot{\ell}_{p,F}(\theta,F|y,x)[F_2-F_1]. \end{aligned}$$

Here, with $\dot{F}(t) = dF(t)/dt$, $\dot{\ell}_{p,\theta} = (\dot{\ell}_{p,\beta},\dot{\ell}_{p,\lambda})^{\intercal}$, $\mathbf{e}_j$ be the d-vector with the j-th component being 1 and others being zeros,

$$\dot{\ell}_{p,\beta}(\theta,F|y,x) = \sum_{j=1}^{r}\frac{\dot{F}(\beta^{\intercal}x+\lambda_j)\big(\gamma_j-\dot{F}(\beta^{\intercal}x+\lambda_j)\big)}{F(\beta^{\intercal}x+\lambda_j)\big(1-F(\beta^{\intercal}x+\lambda_j)\big)}x,$$

$$\dot{\ell}_{p,\lambda}(\theta,F|y,x) = \sum_{j=1}^{r}\frac{\dot{F}(\beta^{\intercal}x+\lambda_j)\big(\gamma_j-\dot{F}(\beta^{\intercal}x+\lambda_j)\big)}{F(\beta^{\intercal}x+\lambda_j)\big(1-F(\beta^{\intercal}x+\lambda_j)\big)}e_j,$$

$$\dot{\ell}_{p,F}(\theta,F|y,x)[h] = \sum_{j=1}^{r}\frac{(\gamma_j-1)h(\beta^{\intercal}x+\lambda_j)}{F(\beta^{\intercal}x+\lambda_j)\big(1-F(\beta^{\intercal}x+\lambda_j)\big)}.$$

By our model specification and (C1)–(C4), both $\dot{\ell}_{p,\theta}(\theta,F|y,x)$ and $\dot{\ell}_{p,F}(\theta,F|y,x)[h]$ have finite second moment. Consequently, by Taylor expansion and Hölder's inequality, there are constants $0 < C_j < \infty$ ($j = 1,2$), such that

$$\begin{aligned} &\|q(Y,X|\theta_1,F_1) - q(Y,X|\theta_2,F_2)\|_{L_1(P)}\\ &\leq C_1\|\theta_1-\theta_2\| + C_2\|F_1-F_2\|_{L_2(P)},\\ &\forall\ (\theta_1,F_1),(\theta_2,F_2) \in (\Theta,\mathcal{F}). \end{aligned}$$

So,

$$N_{[\,]}(\epsilon,\mathcal{Q},||\cdot||_{L_1(P)})$$
$$\leq N_{[\,]}(\frac{\epsilon}{2C_1},\Theta,\|\cdot\|)\times N_{[\,]}(\frac{\epsilon}{2C_2},\mathcal{F},||\cdot||_{L_2(P)}),$$
$$\forall\ \epsilon>0.$$

By (C1), $N_{[\,]}(\epsilon/(2C_1),\Theta,\|\cdot\|)=O(1/\epsilon^d)$, with $d=dim(\Theta)$. Since $\mathcal{F}$ is a collection of bounded monotone functions, by Theorem 2.7.5 in van der Vaart and Wellner (1996, p. 159), for some constant $0<C<\infty$, and for all $r\geq 1$,

$$N_{[\,]}(\epsilon,\mathcal{F},||\cdot||_{L_r(P)})\leq \exp\{\frac{C}{\epsilon}\},\quad \forall\ \epsilon>0.$$

Thus, for some generic constant $0<C<\infty$,

$$N_{[\,]}(\epsilon,\mathcal{Q},||\cdot||_{L_1(P)})$$
$$\leq N_{[\,]}(\frac{\epsilon}{2C_1},\Theta,\|\cdot\|)\times N_{[\,]}(\frac{\epsilon}{2C_2},\mathcal{F},L_2(P))$$
$$\leq \frac{C}{\epsilon^d}\exp\{\frac{C}{\epsilon}\}<\infty,\quad \forall\ \epsilon>0,$$

and so by Theorem 2.4.1 in van der Vaart and Wellner (1996, p. 122), $\mathcal{Q}$ is a Glivenko-Cantelli class with respect to P.

A.2 Proof of Theorem 2

The proof is similar to that of Lemma 1 in Yuan, Yin and Tan (2020) and is omitted.

A.3 Proof of Theorem 3

By Theorem 2, $\|\hat\theta-\theta_0\|=O_p(n^{-1/3})$, so we only need to consider the maximization of $F\in\mathcal{F}_n=\{F:F\in\mathcal{F},\|F-F_0\|\leq Cn^{-1/3}\}$, for some constant $0<C<\infty$. Let $\dot\ell_{p,\theta}(\theta,F|D_n)=\partial\ell_p(\theta,F|D_n)/\partial\theta$ and $\dot\ell_{p,F}(\theta,F|D_n)[h]$ be the Hadamard derivative of $\ell_p(\theta,F|D_n)$ with respect to F in the direction h. Since $(\hat\theta,\hat F)$ is the MLE, by Theorem 2 and (C6) we have

$$\begin{aligned}\hat F&=\arg\max_{F\in\mathcal{F}_n}\frac{1}{n}\ell_p(\hat\theta,F|D_n)\\
&=\arg\max_{F\in\mathcal{F}_n}\Big(\frac{1}{n}\ell_p(\theta_0,F_0|D_n)\\
&\quad+\frac{1}{n}\dot\ell_{p,F}(\theta_0,F_0|D_n)[F-F_0]+\frac{1}{n}\dot\ell_{p,\theta}(\theta_0,F_0|D_n)(\hat\theta-\theta_0)\\
&\quad+O_p\big(\|F-F_0\|^2\big)+O_p\big(\|F-F_0\|\|\hat\theta-\theta_0\|\big)\\
&\quad+O_p\big(\|\hat\theta-\theta_0\|^2\big)\Big)\\
&=\arg\max_{F\in\mathcal{F}_n}\Big(\frac{1}{n}\ell_p(\theta_0,F_0|D_n)\\
&\quad+\frac{1}{n}\dot\ell_{p,F}(\theta_0,F_0|D_n)[F-F_0]\\
&\quad+O_p\big(\|F-F_0\|^2\big)+O_p\big(\|F-F_0\|\|\hat\theta-\theta_0\|\big)\Big)\\
&=\arg\max_{F\in\mathcal{F}_n}\Big(\frac{1}{n}\ell_p(\theta_0,F_0|D_n)\\
&\quad+\frac{1}{n}\dot\ell_{p,F}(\theta_0,F_0|D_n)[F-F_0]+O_p\big(\|F-F_0\|^2\big)\\
&\quad+o_p\big(n^{-2/3}\big)\Big)\\
&=\arg\max_{F\in\mathcal{F}_n}\Big(\frac{1}{n}\ell_p(\theta_0,F|D_n)+o_p\big(n^{-2/3}\big)\Big).\end{aligned}$$

So we have

$$\hat F=\arg\max_{F\in\mathcal{F}_n}\frac{1}{n}\ell(_p\theta_0,F|D_n)+o_p(n^{-2/3}).$$

Denote $\tilde F$ be the minima of the first term above, then $\hat F-F_0=\tilde F-F_0+o_p(n^{-2/3})$. Since $||\hat F-F_0||=O_p(n^{-1/3})$ and $||\tilde F-F_0||=O_p(n^{-1/3})$, thus $n^{1/3}(\hat F(s)-F_0(s))=n^{1/3}(\tilde F(s)-F_0(s))+o_p(n^{-1/3})$. Consequently, $n^{1/3}(\hat F(s)-F_0(s))$ and $n^{1/3}(\tilde F(s)-F_0(s))$ have the same weak limit, and so in this sense we write $\hat F\sim\tilde F$. Let $s_{ij}=\beta_0^{\intercal}x_i+\lambda_{0,j}$, and by the Proposition and the above argument to identify $\hat F$ as

$$\begin{aligned}\hat F&=\arg\max_{F\in\mathcal{F}}\frac{1}{n}\ell_p(\hat\theta,F|D_n)\sim\arg\max_{F\in\mathcal{F}}\frac{1}{n}\ell_p(\theta_0,F|D_n)\\
&=\arg\min_{F\in\mathcal{F}}\sum_{i=1}^n\sum_{j=1}^r\big(F(s_{ij})-t_{ij}\big)^2\\
&=\arg\min_{F\in\mathcal{F}}\sum_{i=1}^n\sum_{j=1}^{r-1}\big(F(s_{ij})-t_{ij}\big)^2.\end{aligned}$$

The last step above holds as $t_{ir}=1=F(\infty)=F(s_{ir})$.

Below for simple of notation, we just write $\hat F=\tilde F$. Recall the definition of P and P_n in the proof of Theorem 1. For the case $r=2$, as in Example 3.2.15 in van der Vaart and Wellner (1996, p. 298), the minima in the above can re-written as a form corresponding to the following. Here the only difference is that we have $n(r-1)$ summands instead of n, and the z_{ij}'s are independent across i but not so across j. Note that t_{ij} iid $I(Y\leq j)$. Let $S=\beta_0^{\intercal}X$, then S_{ij} iid $S+\lambda_{0,j}$. Let P_n be the empirical distribution of $\{(Y_i,\beta_0^{\intercal}X_i):i=1,...,n\}$, and P be the distribution of (Y,S). Define

$$\begin{aligned}V_{n,r}(s)&=\frac{1}{n(r-1)}\sum_{i=1}^n\sum_{j=1}^{r-1}t_{ij}I(s_{ij}<s)\\
&=P_n\Big(\frac{1}{r-1}\sum_{j=1}^{r-1}I(Y\leq j)I(S+\lambda_{0,j}<s)\Big)\\
G_{n,r}(s)&=\frac{1}{n(r-1)}\sum_{i=1}^n\sum_{j=1}^{r-1}I(s_{ij}<s)\\
&=P_n\Big(\frac{1}{r-1}\sum_{j=1}^{r-1}I(S+\lambda_{0,j}<s)\Big).\end{aligned}$$

Then, for $a \in R$,

$$\hat{F}(s) \le a \text{ if and only if } \arg\min_t \{V_{n,r}(t) - aG_{n,r}(t)\} \ge s.$$

Thus for any fixed $x \in R$, with $a = F_0(s) + xn^{-1/3}$,

$$\Big\{n^{1/3}\big(\hat{F}(s) - F_0(s)\big) \le x\Big\} \text{ iff } \Big\{\arg\min_t \big(V_{n,r}(t) - aG_{n,r}(t)\big) - s \ge 0\Big\}.$$

To evaluate the distribution of $n^{1/3}\big(\hat{F}(s) - F_0(s)\big)$, we need to compute the probability of the event $\big\{n^{1/3}\big(\hat{F}(s) - F_0(s)\big) \le x\big\}$ for each $x \in R$. Take $a = F_0(s) + xn^{-1/3}$, then

$$\begin{aligned} &\big\{n^{1/3}\big(\hat{F}(s) - F_0(s)\big) \le x\big\} \\ &= \{\hat{F}(s) \le a\} \\ &= \{n^{1/3}\Big(\arg\min_t \{V_{n,r}(t) - aG_{n,r}(t)\} - s\Big) \ge 0\}, \end{aligned}$$

and by the change of variable $t \mapsto s + n^{-1/3}h$,

$$\begin{aligned} &n^{1/3}\Big(\arg\min_t \{V_{n,r}(t) - aG_{n,r}(t)\} - s\Big) \\ &= \arg\min_h \{V_{n,r}(s + n^{-1/3}h) - \big(F_0(s) + xn^{-1/3}\big) \\ &\quad \times G_{n,r}(s + n^{-1/3}h)\} \\ &= \arg\min_h \{P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(Y \le j) I(S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - \big(F_0(s) + xn^{-1/3}\big) \\ &\quad \times P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(S + \lambda_{0,j} \le s + n^{-1/3}h)\big]\} \\ &= \arg\min_h \{ \\ &\quad P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - xn^{-1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad \} \\ &= \arg\min_h \{ \\ &\quad P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad + P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) I(S + \lambda_{0,j} \le s)\big] \\ &\quad - xn^{-1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - xn^{-1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(S + \lambda_{0,j} \le s)\big] \\ &\quad \} \\ &= \arg\min_h \{ \\ &\quad P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - xn^{-1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad \} \\ &= \arg\min_h \{ \\ &\quad n^{2/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - xn^{1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad \} \\ &= \arg\min_h \{ \\ &\quad n^{2/3}(P_n - P)\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad + n^{2/3} P\big[\frac{1}{r-1}\sum_{j=1}^{r-1} \big(I(Y \le j) - F_0(s)\big) \\ &\quad \times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad - xn^{1/3} P_n\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\quad \} \\ &:= \arg\min_h \{B_{1,n}(h) + B_{2,n}(h) + B_{3,n}(h)\}. \end{aligned}$$

(A.1)

In the above it is understood that if $h < 0$, $(s, s + hn^{-1/3}]$ means $(s + hn^{-1/3}, s]$.

Let $g(\cdot)$ be the density function of S, then

$$\begin{aligned} B_{3,n}(h) &\sim -xn^{1/3} P\big[\frac{1}{r-1}\sum_{j=1}^{r-1} I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big] \\ &\sim -xn^{1/3} \frac{1}{r-1}\sum_{j=1}^{r-1} g(s - \lambda_{0,j}) n^{-1/3}h \end{aligned}$$

$$= -x\frac{1}{r-1}\sum_{j=1}^{r-1} g(s-\lambda_{0,j})h$$
$$:= -x\bar{g}(s)h.$$

Note that $E[I(Y \le j)|S=t] = F_0(t+\lambda_{0,j})$, so

$$B_{2,n}(h) = n^{2/3}\int \frac{1}{r-1}\sum_{j=1}^{r-1} E\big[\big(I(Y\le j) - F_0(s)\big)$$
$$\times I(s \le S + \lambda_{0,j} \le s + n^{-1/3}h)\big|S=t\big]g(t)dt$$
$$= n^{2/3}\frac{1}{r-1}\sum_{j=1}^{r-1}\int_{s-\lambda_{0,j}}^{s-\lambda_{0,j}+n^{-1/3}} [F_0(t+\lambda_{0,j}) - F_0(s)]g(t)dt$$
$$\sim n^{2/3}\frac{1}{r-1}\sum_{j=1}^{r-1}\dot{F}_0(s)g(s-\lambda_{0,j})\int_{s-\lambda_{0,j}}^{s-\lambda_{0,j}+n^{-1/3}h}(y-s)dy$$
$$= n^{2/3}\dot{F}_0(s)\frac{1}{r-1}\sum_{j=1}^{r-1} g(s-\lambda_{0,j})n^{-2/3}\frac{1}{2}h^2 = \frac{1}{2}\dot{F}_0(s)\bar{g}(s)h^2.$$

Now we evaluate $B_{1,n}(h)$. For $h<0$, the notation $I(s \le S \le s + hn^{-1/3})$ means $I(s + hn^{-1/3} \le S \le s)$. Let $S_i = \beta^{\mathsf{T}} X_i$,

$$f_{n,h}(Y_i, S_i) = n^{1/6}\frac{1}{r-1}\sum_{j=1}^{r-1}\big(I(Y_i \le j) - F_0(s)\big)$$
$$\times I(s \le S_i + \lambda_{0,j} \le s + hn^{-1/3}).$$

Then

$$B_{1,n}(h) = n^{-1/2}\sum_{i=1}^{n}\big[f_{n,h}(Y_i,S_i) - Pf_{n,h}(Y_i,S_i)\big].$$

Let $\mathcal{F}_n = \{f_{n,h}(\cdot,\cdot) : |h| \le K\}$, for some $K>0$. $\mathcal{F}_n$ has an envelope function $F_n(y,S) = n^{1/6}(r-1)^{-1}\sum_{j=1}^{r-1} I(s - Kn^{-1/3} \le S + \lambda_{0,j} \le s + Kn^{-1/3})$, with some $0<C<\infty$. It is apparent that

$$PF_n^2(Y,S) = O(1), \quad P\Big[F_n^2 I(F_n > \eta\sqrt{n})\Big] \to 0, \quad \forall\ \eta > 0.$$

Also, for fixed (b,h) and large n, $I(s + (b\wedge h)n^{-1/3} \le S + \lambda_{0,j}, S + \lambda_{0,k} \le s + (b \vee h)n^{-1/3}) = 0$ for $j \ne k$, so some $0 < C < \infty$,

$$P\big(f_{n,b} - f_{n,h}\big)^2 = n^{1/3}\frac{1}{(r-1)^2}\sum_{j=1}^{r-1} E\Big[\big(I(Y\le j) - F_0(s)\big)^2$$
$$\times I\big(s + (b\wedge h)n^{-1/3} \le S + \lambda_{0,j} \le s + (b\vee h)n^{-1/3}\big)\Big]$$
$$\sim Cn^{1/3}\frac{1}{(r-1)^2}\sum_{j=1}^{r-1}\int_{s-\lambda_{0.j}+(b\wedge h)n^{-1/3}}^{s-\lambda_{0,j}+(b\vee h)n^{-1/3}h} g(y)dy$$
$$\sim O(1)|b-h|,$$

and so for any totally bounded semi-metric $\rho(\cdot,\cdot)$ on $[-K,K]$,

$$\sup_{(b,h):\rho(r,h)\le\delta_n} P\big(f_{n,b} - f_{n,h}\big)^2 \sim O(1)|b-h| \to 0, \quad \forall\ \delta_n \to 0.$$

Thus the three conditions in (2.11.21) of Theorem 2.11.22 of van der Vaart and Wellner (1996, p. 220) are satisfied.

Let $\mathcal{F}_{n,\delta} = \{f_{n,b} - f_{n,h} : \rho(b,h) \le \delta\}$. Then $\mathcal{F}_{n,\delta}$ and $\mathcal{F}^2_{n,\delta}$ are P-measurable in the sense of definition 2.33 in van der Vaart and Wellner (1996, p. 110). Also, for fixed h,

$$Pf_{n,h} = n^{1/6}\frac{1}{r-1}\sum_{j=1}^{r-1} E\Big[\big(I(Y\le j) - F_0(s)\big)$$
$$\times I(s \le S + \lambda_{0,j} \le s + hn^{-1/3})\Big]$$
$$= n^{1/6}\frac{1}{r-1}\sum_{j=1}^{r-1}\int_{s-\lambda_{0,j}}^{s-\lambda_{0,j}+hn^{-1/3}}\big(F_0(t+\lambda_{0,j})$$
$$- F_0(s)\big)g(t)dt$$
$$\sim n^{1/6}\dot{F}_0(s)\frac{1}{r-1}\sum_{j=1}^{r-1} g(s-\lambda_{0,j})n^{-2/3}\frac{h^2}{2} \to 0,$$

and for $b>0$ and $h>0$,

$$Pf_{n,b}f_{n,h} = n^{1/3}\frac{1}{(r-1)^2}\sum_{j=1}^{r-1} E\Big[\big(I(Y\le j) - F_0(s)\big)^2$$
$$\times I\big(s \le S + \lambda_{0,j} \le s + (b\wedge h)n^{-1/3}\big)\Big]$$
$$= n^{1/3}\frac{1}{(r-1)^2}\sum_{j=1}^{r-1}\int_{s-\lambda_{0,j}}^{s-\lambda_{0,j}+(b\wedge h)n^{-1/3}}$$
$$E\Big[\big(I(Y\le j) - F_0(s)\big)^2\Big|S=t\Big]g(t)dt$$
$$\sim n^{1/3}\frac{1}{(r-1)^2}$$
$$\times\sum_{j=1}^{r-1} E\Big[\big(I(Y\le j) - F_0(s)\big)^2\Big|S = s-\lambda_{0,j}\Big]$$
$$\times g(s-\lambda_{0,j})(b\wedge h)n^{-1/3}$$
$$= \frac{1}{(r-1)^2}\sum_{j=1}^{r-1}$$
$$\times E\Big[\big(I(Y\le j) - F_0(s)\big)^2\Big|S = s-\lambda_{0,j}\Big]$$
$$\times g(s-\lambda_{0,j})(b\wedge h)$$
$$:= \eta^2(s)(b\wedge h).$$

It can be checked that the above also holds for any b and h. So that as $n\to\infty$, $\big(Pf_{n,r}f_{n,h} - Pf_{n,r}Pf_{n,h}\big) \to \eta^2(s)(b\wedge h)$. Note that $E[\mathbb{B}(s) - \mathbb{B}(h)]^2 = |s-h|$, where $\mathbb{B}(\cdot)$ is the two-sided Brownian motion process, originating from zero. This gives that $Cov(f_{n,r}, f_{n,h}) = P\big(f_{n,r}f_{n,h}\big) - P(f_{n,r})P(f_{n,h})$ converges to the covariance function of the process $\eta(s)\mathbb{B}(\cdot)$.

For a probability measure Q, let $N(\epsilon, \mathcal{F}_n, L_2(Q))$ be the number of ϵ-balls needed to cover $\mathcal{F}_n$ under the metric of $L_2(Q)$. It is easy to see that $\sup_Q N(\epsilon||F_n||_{Q,2}, \mathcal{F}_n, L_2(Q)) = O(1/\epsilon)$, where the supreme is over all probability measures. Thus, $\forall\ \eta_n \to 0$,

$$\begin{aligned}&\sup_Q \int_0^{\eta_n} \sqrt{N(\epsilon||F_n||_{Q,2}, \mathcal{F}_n, L_2(Q))}d\epsilon \\ &\sim \int_0^{\eta_n} \sqrt{-\log \epsilon}d\epsilon = 2\int_{\sqrt{-\log \eta_n}}^{\infty} xe^{-x^2}dx \to 0,\end{aligned}$$

and so by Theorem 2.11.22 in van der Vaart and Wellner (1996, p. 220),

$$B_{1,n}(\cdot) \xrightarrow{D} \eta(\cdot)\mathbb{B}(\cdot), \qquad \text{in } l^\infty([-K, K]),$$

where $l^\infty([-K, K])$ is the space of all bounded real functions on $[-K, K]$ equipped with the supreme metric.

Now collecting results from (A.1), we have that the event $\Big\{n^{1/3}\big(\hat{F}(s) - F_0(s)\big) \le x\Big\}$ is asymptotically equivalent to

$$\begin{aligned}&\Big\{\arg\min_h \big\{\eta(s)\mathbb{B}(h) + \frac{1}{2}\dot{F}_0(s)\bar{g}(s)h^2 - x\bar{g}(s)h\big\} \ge 0\Big\} \\ &= \Big\{\arg\max_h \big\{-\frac{\eta(s)}{\bar{g}(s)}\mathbb{B}(h) - \frac{1}{2}\dot{F}_0(s)h^2 + xh\big\} \ge 0\Big\} \\ &= \Big\{\arg\max_h \big\{\frac{\eta(s)}{\bar{g}(s)}\mathbb{B}(h) - \frac{1}{2}\dot{F}_0(s)h^2 + xh\big\} \ge 0\Big\}.\end{aligned}$$

In the above we used the fact that $-\mathbb{B}(h) \overset{D}{=} \mathbb{B}(h)$. Using problem 3.2.5 in van der Vaart and Wellner (1996, p. 308), the above is re-written as

$$\begin{aligned}&\Big\{\Big(\frac{4\eta^2(s)\dot{F}_0(s)}{\bar{g}^2(s)}\Big)^{1/3} \arg\max_h \big\{\mathbb{B}(h) - h^2\big\} + x \ge 0\Big\} \\ &= \Big\{\Big(\frac{4\eta^2(s)\dot{F}_0(s)}{\bar{g}^2(s)}\Big)^{1/3} \arg\min_h \big\{-\mathbb{B}(h) + h^2\big\} \ge -x\Big\} \\ &= \Big\{\Big(\frac{4\eta^2(s)\dot{F}_0(s)}{\bar{g}^2(s)}\Big)^{1/3} \arg\min_h \big\{\mathbb{B}(h) + h^2\big\} \ge -x\Big\} \\ &= \Big\{\Big(\frac{4\eta^2(s)\dot{F}_0(s)}{\bar{g}^2(s)}\Big)^{1/3} \arg\min_h \big\{\mathbb{B}(h) + h^2\big\} \le x\Big\}, \quad \forall\ s,\ x.\end{aligned}$$

In the above we used the fact that $-\mathbb{B}(\cdot)$ and $\mathbb{B}(\cdot)$ have the same distribution, and that $W = \arg\min_h \big\{\mathbb{B}(h) + h^2\big\}$ is symmetrically distributed about 0, thus $P\big(CW \ge -b\big) = P\big(CW \le b\big)$. This gives the desired result for $h \in [-K, K]$. Below we prove that the result is actually true on R. Let $\hat{h}_n = \arg\min_h\{V_n(h) - aU_n(h)\} - t$. We need to show that $\hat{h}_n$ is bounded in probability. Thus for large n, $\hat{h}_n$ will be in $[-K, K]$ for some $0 < K < \infty$ in probability, and so the desired result is true on R. For this, we only need to show $n^{1/3}d(\hat{h}_n, \hat{h}) = O_p(1)$ for some distance $d(\cdot, \cdot)$. The method is similar to that in the proof of the Lemma and is omitted.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their helpful suggestions which improved the quality of our paper, and the Editor for careful reading of the manuscript that improved its presentation. This work was partially supported by the National Natural Science Foundation of China (81872710).

Received 5 November 2019

REFERENCES

Adeyemi, R., Zewotir, T., and Ramroop, S. (2016). Semiparametric multinomial ordinal model to analyze spatial patterns of child birth weight in Nigeria. *International Journal of Environmental Research and Public Health* **13** 1145. doi:10.3390/ijerph13111145.

Agresti, A. (1999). Modelling ordered categorical data: recent advances and future challenges. *Statistics in Medicine* **18** 2191–2207.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley, second edition. MR2742515

Amini, A. A., Chen, A., Bickel, P., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics* **41**(4) 2097–2122. MR3127859

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. B* **46** 1–30. MR0745211

Bickel, P. J. and Fan, J. (1996). Some problems on the estimation of unimodal densities. *Statistica Sinica* **6** 23–45. MR1379047

Best, M. J. and Chakravarti, N. (1990). Active set algorithms for isotonic regression; a unifying framework. *Mathematical Programming* **47** 425–439. MR1068274

Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Annals of Statistics* **17** 1532–1549. MR1026298

Chen, M. H., Dey, D. K., and Shao, Q. M. (1999). A new skewed link model for dichotomous quantal response data. *J. Am. Stat. Assoc.* **94** 1172–1186. MR1731481

Cox, D. R. and Reid, N. (2004). A note on pseudo-likelihood constructed from marginal densities. *Biometrika* **91**(3) 729–737. MR2090633

Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51**(3) 765–782. MR0712369

Crouchley, R. (1995). A random-effects model for ordered categorical data. *Journal of the American Statistical Association* **90** 489–498.

Das, K., Roy, S., and Chattopadhyay, A. K. (2016). Analysis of ordinal longitudinal data using semi-parametric mixed models. *Journal of Statistical Research* **48–50**(1) 15–33.

Fahrmeir, L. and Pritscher, L. (1996). Regression analysis of forest damage by marginal models for correlated ordinal responses. *Journal of Environmental and Ecological Statistics* **3** 257–268.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association* **74** 537–552. MR0548257

Grenander, U. (1956). On the theory of mortality measurement. Part II. *Skand. Aktuarietidskr* **39** 125–153. MR0093415

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*, Cambridge University Press. MR3445293

Groeneboom, P. and Hendrickx, K. (2018). Current status linear regression. *Annals of Statistics* **46** 1415–1444. MR3819105

Harville, D. A. and Mee, R. W. (1984). A mixed model procedure for analyzing ordered categorical data. *Biometrics* **40** 393–408. MR0763742

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Society* **91** 1024–1036.

Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50** 933–944.

Huang, J. (2002). A note on estimating a partly linear model under monotonicity constraint. *Journal of Statistical Planning and Inference* **107** 343–351. MR1927773

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* D. Lin and T. Fleming (eds), 123–169. Springer-Verlag, New York.

Kauermann, G., Tutz, G. (2003). Semi- and nonparametric modeling of ordinal data. *Journal of Computational and Graphical Statistics* **21**(1) 176–196. MR1965214

Kazembe, L. (2009). A semiparametric sequential ordinal model with applications to analyse first birth intervals. *Australian Journal of Statistics* **38**(2) 83–99.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotic minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85. MR0397974

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrika* **61**(2) 387–421. MR1209737

Kottas, A., Müller, P., and Quintana, F. (2005). Nonparametric Bayesian Modelling for Multivariate Ordinal Data. *Journal of Computational and Graphical Statistics* **14**(3) 610–625. MR2170204

Laara, E. and Matthews, J. (1985). The equivalence of two models for ordinal data. *Biometrika* **72** 206–207.

Lo, S. H. (1986). Estimation of a unimodal distribution function. *Annals of Statistics* **14** 1132–1138. MR0856810

Mair, P., Hornik, K., de Leeuw, J. (2009). Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of statistical software* **32**(5) 1–24.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, B* **42** 109–142. MR0583347

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London, second edition. MR3223057

Murphy, S. A., van der Vaart, A. W., and Wellner, J. A. (1999). Current status regression. *Mathematical Methods of Statistics* **8**(3) 407–425. MR1735473

Patilea, V. (2001). Convex models, MLE and misspecification. *Annals of Statistics* **29** 94–123. MR1833960

Pfanzagl, J. (1990). Large deviations probabilities for certain nonparametric maximum likelihood estimators. *Annals of Statistics* **18** 1868–1877. MR1074441

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankya (A)* **31** 23–36. MR0267677

Prakasa Rao, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics* **41** 507–519. MR0260133

Qin, J., Garcia, T. P., Ma, Y., Tang, M.-X., Marder K., and Wang, Y. (2014). Combining isotonic regression and EM algorithm to predict risk under monotonicity constraint. *Annals of Applied Statistics* **8** 1182–1208. MR3262550

Qu, Y. and Tan, M. (1998). Analysis of clustered ordinal data with sub-clusters via a Bayesian hierarchical model. *Communications in Statistics Series A: Theory & Methods* **27** 1461–1476.

Robertson, T., Wright, F. T., and Dykstra, R. (1988). *Order Restricted Statistical Inference*. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore. MR0961262

Shmueli, G. (2010). To explain or to predict? *Statistical Science* **25**(3) 289–310. MR2791669

Stout, Q. F. (2008). Unimodal regression via prefix isotonic regression. *Computational Statistics and Data Analysis* **53** 289–297. MR2649085

Sutradhar, B. C. (2017). Semi-parametric dynamic models for longitudinal ordinal categorical data. *Sankhya A: The Indian Journal of Statistics*. doi:10.1007/s13171-017-0100-z. MR3772092

Tan, M., Qu, Y., Mascha, E., and Schubert, A. (1999). A Bayesian hierarchical model for multi-level repeated ordinal data: analysis of oral practice examinations in a large anaesthesiology training programme. *Statistics in Medicine* **18** 1983–1992.

Turnball, B. C. and Ghosh, S. K. (2016). Unimodal density estimation using Bernstein polynomilas. Manuscript. MR3139345

Tutz, G. (2003). Generalized semiparametrically structured ordinal models. *Biometrics* **59** 263–273. MR1987393

van der Vaart, A. (2002). Semiparametric Statistics. In Part III, *Lectures on Probability Theory and Statistics*, Springer. MR1915446

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*, Springer. MR1385671

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**(3) 439–447. MR0375592

Williamson, J. M., Kim, K., and Lipsitz, S. R. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association* **90** 1432–1437. MR1379487

Xia, Y. and Gou, J. (2016). Bayesian semiparametric analysis for latent variable models with mixed continuous and ordinal outcomes. *Journal of the Korean Statistical Society* **45**(3) 451–465. MR3527659

Yuan, A., Piao, J., Ning, J., and Qin, J. (2020). Semiparametric isotonic regression model and estimation for group testing data. *Canadian Journal of Statistics*, In press.

Yuan, A., Yin, A., Tan, M. T. (2020). Semiparametric subgroup causal inference on treatment difference. In revision to a Journal.

Ao Yuan
Department of Biostatistics
Bioinformatics and Biomathematics
Georgetown University
Washington DC 20057
USA
E-mail address: ay312@georgetown.edu

Chongyang Duan
Department of Biostatistics
School of Public Health
Southern Medical University
Guangzhou 510515
China
E-mail address: donyduang@126.com

Ming T. Tan
Department of Biostatistics
Bioinformatics and Biomathematics
Georgetown University
Washington DC 20057
USA
E-mail address: mtt34@georgetown.edu