Published in final edited form as:

Comput Stat Data Anal. 2019 June ; 134: 86–210. doi:10.1016/j.csda.2018.11.001.

A Smooth Nonparametric Approach to Determining Cut-Points of A Continuous Scale

Zhiping Qiu^{a,b}, Limin Peng^{a,*}, Amita Manatunga^a, Ying Guo^a

^aDepartment of Biostatistics and Bioinformatics, Emory University, Atlanta, U.S.A.

^bSchool of Mathematical Sciences, Huaqiao University, Quanzhou, China

Abstract

The problem of determining cut-points of a continuous scale according to an establish categorical scale is often encountered in practice for the purposes such as making diagnosis or treatment recommendation, determining study eligibility, or facilitating interpretations. A general analytic framework was recently proposed for assessing optimal cut-points defined based on some pre-specified criteria. However, the implementation of the existing nonparametric estimators under this framework and the associated inferences can be computationally intensive when more than a few cut-points need to be determined. To address this important issue, a smoothing-based modification of the current method is proposed and is found to substantially improve the computational speed as well as the asymptotic convergence rate. Moreover, a plug-in type variance estimation procedure is developed to further facilitate the computation. Extensive simulation studies confirm the theoretical results and demonstrate the computational benefits of the proposed method. The practical utility of the new approach is illustrated by an application to a mental health study.

Keywords

Agreement; Association; Cut-point; Nonparametric; Smoothing objective function

1. Introduction

In many biomedical studies, often, it is of interest to determine the cut-points or discrimination values for categorizing a continuous diagnostic test, for example, to make treatment recommendations, determine study eligibility, or facilitate interpretations. For example, in the Diabetes and Depression Study [1], the 20-item self-rating Zung Depression scale [2] was administered to aid in the assessment of depression severity in diabetes subjects. Compared with clinician-administered instruments, Zung has the advantage of being less time consuming and more cost effective; however there is no well-established cut-points of Zung to reflect the degree of severity of depression (e.g. no/mild depression,

*Corresponding to: Department of Biostatistics and Bioinformatics, 1518 Clifton Road, NE, Atlanta, GA 30322, USA. Tel: +1 404 727 7701. Fax: +1 404 727 1370, lpeng@emory.edu (Limin Peng).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

moderate depression and severe depression). Having such cut-points for Zung will facilitate identifying depression subjects for large populations.

Various analytic strategies have been used in practice to categorize a continuous scale, most of which are limited to a special case of identifying a single cut-point. Some intuitive approaches include considering arbitrary cut-points; specifying cut-points as sample quantiles (e.g. median) or according to clinicians' experience [3]; using cut-points that yield disease rates consistent with a known population disease prevalence, or the highest proportions of correct classification based on a gold standard [4, 5]. Another popular approach is to utilize the receiver operating characteristic (ROC) curve [6] in conjunction with various accuracy measures, such as Youden index [7, 8, 9, 10, 11, 12, 13], concordance probability methods [14] and the point closest-to-(0,1) corner method [15], among others. Dong et al. [16] proposed to estimate cut-points based on a new accuracy measure, maximum absolute determinant (MADET), which can handle a ordinal scale with more than two levels and includes Yondex's index as a special case. Other types of methods were proposed in literature, for example, determining a single cut-point by minimizing the variance estimator of the estimated group prevalences under the mixture normal assumption [17, 18, 19], or by maximizing a distance measure between the differentiated two groups [20, 21, 22, 23, 24].

More recently, Peng et al. [25] proposed a general nonparametric analytic framework for defining and assessing optimal cut-points based on a pre-specified criterion, which can be an association or agreement measure, or an accuracy measure attached to ROC (e.g. Youden index). In particular, Peng et al. [25]'s framework broadly covers the cases where one or more than one cut-points need to be determined, without imposing parametric assumptions. More importantly, Peng et al. [25] uncovered that standard asymptotic theory, such as root- n asymptotic normality, and routine inferential procedures (e.g. traditional bootstrapping) may not be valid in their framework. Accordingly, Peng et al. [25] derived rigorous nonstandard theory and develop appropriate inferential procedures for their proposed cut-point identification.

Though Peng et al. [25]'s framework is broad and theoretically sound, the implementation of their method may involve heavy computation, which can limit its practical applications. The issue is that Peng et al. [25] proposed to evaluate the empirical objective function at all possible sets of cut-points, the number of which is roughly of order $O(n^{L-1})$, where n denotes the sample size, and $L - 1$ corresponds to the number of cut-points that need to be determined. This means the computation loan increases significantly when n becomes large. Additionally, the currently available inference procedure is built upon resampling, which further increase the computational demand.

To address the obstacle discussed above, in this paper we propose to apply the smooth approximation technique to the estimation of optimal cut-points under the framework of Peng et al. [25]. In literature, the smooth approximation techniques have been used in various context [26, 27, 28, 29, 30, for example]. In this work, the basic idea is to replace the non-smooth objective function of Peng et al. [25] by a smooth surrogate, which is asymptotic equivalent to the original one. As demonstrated by our numerical studies, this

can lead to substantial savings in computation time (e.g. ~ 25 times faster when $n = 300$). Moreover, we provide rigorous theoretical justifications for the new estimator derived based on the smooth approximation of the objective function. Of note, under certain regularity conditions, we can show that the convergence rate of the new cut-point estimator can be arbitrarily close to root- n with appropriately selected bandwidth and local distribution function, while the original estimator proposed in Peng et al. [25] only achieves the cubic- n convergence rate. This result reflects a rather significant efficiency gain resulted from using the new nonparametric approach. In addition, we derive a plug-in type variance estimator for the estimated cut-points. This is shown to further expedite the computation.

The rest of paper is organized as follows. In Section 2, we first review the framework of Peng et al. [25] for defining and estimating optimal cut-points and then present the proposed nonparametric estimator of the optimal cut-points. In Section 3, we present our asymptotic studies and develop inference procedures, including covariance estimation as well as the selection of the local distribution function and bandwidth parameter involved in the proposed estimation procedure. We conduct extensive numerical studies, including Monte-Carlo simulations and an application to the Diabetes and Depression study, which are presented in Section 4. Concluding comments are given in Section 5.

2. The Proposed Estimation Procedure

2.1. A review of Peng et al. [25]'s framework

Let Y denote an ordinal measurement that takes ordinal values, $1 < \dots < L$, and X denote a continuous measurement with compact support $[x_L, x_U]$. For any vector $\mathbf{b}$, we use $\mathbf{b}^T$ to denote its transpose. Assume that all possible cut-points form a compact parameter space $\Theta = \{(d_1, d_2, \dots, d_{L-1})^T: d_0 = x_L < d_1 < \dots < d_{L-1} < d_L = x_U\}$. For $\vec{d} = (d_1, \dots, d_{L-1})^T \in \Theta$, define

$$\tilde{X}(\vec{d}) = \sum_{k=1}^L kI(d_{k-1} < X \leq d_k), p_{ij}(\vec{d}) = Pr(\tilde{X}(\vec{d}) = i, Y = j) (i = 1, 2, \dots, L, j = 1, 2, \dots, L),$$

and $\mathbf{P}(\vec{d}) = (p_{11}(\vec{d}), \dots, p_{1L}(\vec{d}), \dots, p_{L1}(\vec{d}), \dots, p_{LL}(\vec{d}))^T$. Peng et al. [25] proposed to define the optimal set of cut-points, $\vec{d}_0$, based on a pre-specified criterion function that takes a general form $\vartheta(\mathbf{P}(\vec{d}))$, where $\vartheta(\cdot)$ is a smooth function from $[0, 1]^{L^2}$ to R , i.e.

$$\vec{d}_0 = \arg \max_{\vec{d} \in \Theta} \vartheta(\mathbf{P}(\vec{d})). \quad (1)$$

As illustrated in Peng et al. [25], the criterion function $\vartheta(\mathbf{P}(\vec{d}))$ with different choices of $\vartheta(\cdot)$ can include a wide range of agreement, association and classification accuracy measures, including the weighted kappa coefficient, correct classification rate, Kendall's τ_b , Youden's index for ROC and the maximum absolute determinant (MADET) newly proposed by Dong et al. [16].

Peng et al. [25] further proposed a nonparametric approach for estimating the optimal cut-point $\vec{d}_0$, which is given by

$$\widehat{\vec{d}} = \arg \max_{\vec{d} \in \Theta} \vartheta(\mathbf{P}_n(\vec{d})), \quad (2)$$

where $\mathbf{P}_n(\vec{d}) = (\hat{p}_{11}(\vec{d}), \dots, \hat{p}_{1L}(\vec{d}), \dots, \hat{p}_{L1}(\vec{d}), \dots, \hat{p}_{LL}(\vec{d}))^T$ and

$$\hat{p}_{ij}(\vec{d}) = \frac{1}{n} \sum_{k=1}^n I(\tilde{X}_k(\vec{d}) = i, Y_k = j) = \frac{1}{n} \sum_{k=1}^n I(d_{i-1} < X_k \leq d_i) I(Y_k = j), i, j = 1, 2, \dots, L,$$

(3)

and (X_k, Y_k) , $k = 1, 2, \dots, n$ are i.i.d samples from the population (X, Y) .

Note that the objective function in (2) involves indicator functions $I(d_{i-1} < X_k \leq d_i) = I(X_k \leq d_i) - I(X_k \leq d_{i-1})$ ($i = 1, \dots, L$) and thus is discontinuous with respect to $\vec{d}$. This poses additional challenges for the computation and theory of the maximization problem in (2). The Nelder-Mead method [31] can be applied but may be unstable particularly when the objective function is not concave or asymptotically concave. To ensure locating the true maximizer, Peng et al. [25] adopted an exact search algorithm. That is, they proposed evaluating $\vartheta(\mathbf{P}_n(\vec{d}))$ at all sets of cut-points where this objective function possibly changes its value and then find the one that produces the largest value of the objective function. While this algorithm guarantees finding the exact solution to the maximization problem in (2), it can be computationally intensive because the evaluation of $\vartheta(\mathbf{P}_n(\vec{d}))$ may need to be done at up to $\binom{n}{L-1}$ sets of cut-points. When Youden's index or correct classification rate is adopted as the criterion for defining optimal cut-points, Nakas et al. [32]'s work suggested some simplifications such that the number of candidate sets of cut-points may be reduced to be around $\prod_{l=1}^{L-1} (n_l + n_{l+1})$. Nonetheless, this number can still be very large, leading to prohibitive computation. Furthermore, inferences, such as variance estimation and confidence intervals, were proposed to perform based on subsampling [33], which demands repeating the estimation procedure a number of times. These suggest a practical limitation of Peng et al. [25]'s method in the case where the dataset is of moderate or large sample size and when L increases. For example, according to our simulation studies, given a simulated dataset with $n = 500$ and $L = 3$, the computation time for getting $\widehat{\vec{d}}$ and its standard error is about 16,930 minutes. This issue is further illustrated in Section 4.

2.2. The proposed nonparametric estimator

We propose to facilitate the estimation of $\vec{d}_0$ by approximating the indicator function $I(X_k d_j)$ by a local distribution function $S((d_j - X_k)/h_n)$. Adopting this strategy allows us to obtain a smooth approximation to the objective function $\vartheta(\vec{P}_n(\vec{d}))$ in (2). Here $S(u)$ is a twice-differentiable and non-decreasing function and satisfies the conditions $\lim_{u \rightarrow \infty} S(u) = 1$ and $\lim_{u \rightarrow -\infty} S(u) = 0$, and h_n is a sequence of strictly positive and decreasing numbers satisfying $\lim_{n \rightarrow \infty} h_n = 0$. Clearly, when $d_j > X_k$, $S((d_j - X_k)/h_n) \rightarrow 1$ as $n \rightarrow \infty$; on the other hand, when $d_j < X_k$, $S((d_j - X_k)/h_n) \rightarrow 0$ as $n \rightarrow \infty$. Therefore, $S((d_j - X_k)/h_n) - S((d_{j-1} - X_k)/h_n)$ provides a smooth approximation for the indicator function $I(d_{j-1} < X_k d_j)$ in (2).

Mimicking (2), we propose to estimate $\vec{d}_0$ by the following nonparametric estimator,

$$\hat{\vec{d}}_s = \arg \max_{\vec{d} \in \Theta} \vartheta(\vec{P}_n(\vec{d})), \quad (4)$$

Where $\vec{P}_n(\vec{d}) = (\tilde{p}_{11}(\vec{d}), \dots, \tilde{p}_{1L}(\vec{d}), \dots, \tilde{p}_{L1}(\vec{d}), \dots, \tilde{p}_{LL}(\vec{d}))^T$, and

$$\tilde{p}_{ij}(\vec{d}) = \frac{1}{n} \sum_{k=1}^n \left[S\left(\frac{d_i - X_k}{h_n}\right) - S\left(\frac{d_{i-1} - X_k}{h_n}\right) \right] I(Y_k = j), \quad i, j = 1, 2, \dots, L. \quad (5)$$

We elaborate the technical conditions for the local distribution function $S(\cdot)$ and discuss its practical examples in Section 3.

It is clear that the objective function $\vartheta(\vec{P}_n(\vec{d}))$ is differentiable with respect to $\vec{d}$. Thus, we can efficiently solve the maximization problem in (4) by many existing algorithms such as the Newton-Raphson algorithm and its variants [34]. Some of these methods require specifying an initial value, an inappropriate selection of which may yield a local maximum and hence an inconsistent estimator. To overcome this issue, we adopt a differential evolution algorithm developed by [35], which is a global optimization algorithm to optimize a real-valued function and has been applied in a wide variety of fields [36]. Specifically, we solve the maximization problem in (4) by using the R function *DEoptim()* in the *DEoptim* package.

It is interesting to note that $\tilde{p}_{ij}(\vec{d})$ can be alternatively expressed as

$$\tilde{p}_{ij}(\vec{d}) = \tilde{F}_n(d_i | Y = j) \hat{P}_n(Y = j) - \tilde{F}_n(d_{i-1} | Y = j) \hat{P}_n(Y = j), \quad (6)$$

where $\tilde{F}_n(x|Y=j) = \sum_{k=1}^n S\left(\frac{x-X_k}{h_n}\right) I(Y_k=j) / \sum_{k=1}^n I(Y_k=j)$ and $\hat{P}_n(Y=j) = \frac{1}{n} \sum_{k=1}^n I(Y_k=j)$, which respectively correspond to the smooth Nadaraya-Watson estimator [37] of the conditional distribution function $F(x|Y=j) \doteq P(X \leq x|Y=j)$ and the empirical estimator of $P(Y=j)$. At the same time, we can write $\hat{p}_{ij}(\vec{d}) = \hat{F}_n(d_i|Y=j)\hat{P}_n(Y=j) - \hat{F}_n(d_{i-1}|Y=j)\hat{P}_n(Y=j)$, where $\hat{F}_n(x|Y=j) \doteq \sum_{k=1}^n I(X_k \leq x, Y_k=j) / \sum_{k=1}^n I(Y_k=j)$ is the empirical distribution estimator of $F(x|Y=j)$. As commented in Hansen [37], estimating $F(x|Y=j)$ by $\tilde{F}_n(x|Y=j)$ instead of $\hat{F}_n(x|Y=j)$ can result in variance reductions. Consequently we expect the proposed new nonparametric estimator $\widehat{\vec{d}}_s$ may have improved statistical efficiency as compared to Peng et al. [25]'s estimator. This was confirmed by our theoretical studies in Section 3 and simulation studies in Section 5. Moreover, based on (6), one can further improve the estimation efficiency of the proposed method by replacing $\hat{P}_n(Y=j)$ by Li and Racine [38]'s estimator of $P(Y=j)$, which uses the discrete univariate kernel function. That is, we can estimate the probability $P(Y=j)$ by

$$\hat{P}_n^*(Y=j) = \frac{1}{n} \sum_{i=1}^n l(Y_i, j, \lambda_s),$$

where $l(Y_i, j, \lambda_s)$ is the discrete univariate kernel function defined as follows:

$$l(Y_i, j, \lambda_s) = \begin{cases} 1 - \lambda_s, & \text{if } Y_i = j, \\ \lambda_s/L, & \text{if } Y_i \neq j, \end{cases}$$

and $\lambda_s \in [0, L/(L+1)]$. For presentation simplicity, we shall not pursue the refinement of this kind.

3. Asymptotic Studies and Inference Procedures

3.1. Asymptotic properties of $\widehat{\vec{d}}_s$

In this subsection, we study the asymptotic properties of the proposed estimator $\widehat{\vec{d}}_s$. Let $f(x)$ denote the density function of X with respect to the Lebesgue measure, $f(x|Y=j)$ denote the conditional density function of X given $Y=j$ and $f^{(m)}(x|Y=j)$ ($j=1, \dots, L$) denote the m -th derivative of $f(x|Y=j)$ with respect to x . In addition, $\vartheta'_j(p)$ denotes the first derivative of $\vartheta(p)$ with respect to the j th component of p , and $\vartheta''_{ij}(p)$ denotes the second derivative of $\vartheta(p)$ with respect to the i th and j th components of p ($i, j=1, \dots, L^2$). We assume the following regularity conditions:

(C1) (i) $\vartheta(P(\vec{d}))$ is twice differentiable with respect to $\vec{d}$ and its second-derivative matrix $Q(\vec{d})$ (specified in the Appendix) is bounded and nonsingular; (ii) $\vec{d}_0$ is the unique maximizer of $\vartheta(P(\vec{d}))$.

(C2) (i) $f(x)$ is positive everywhere; (ii) for some integer $r \geq 2$ and $j = 1, \dots, L$, the r -th order derivative of $f^{(r)}(x|Y=j)$ exists and is continuous and bounded.

(C3) $\{h_n, n = 1, 2, \dots\}$ is a sequence of strictly positive real numbers satisfying the following conditions: $h_n \rightarrow 0$, $nh_n \rightarrow \infty$, $(nh_n)^{-1} \log(n) \rightarrow 0$ and $n^{-1}h_n^{-3} \log(1/h_n) \rightarrow 0$ as $n \rightarrow \infty$.

(C4) (i) $S(u)$ is twice differentiable and non-decreasing with respect to u and satisfies the conditions: $\lim_{u \rightarrow \infty} S(u) = 1$, $\lim_{u \rightarrow -\infty} S(u) = 0$; (ii) $S'(u) = S'(-u)$, $\int S'(u) du = 1$, $|S'(\cdot)|$ and $|S''(\cdot)|$ are uniformly bounded and the following integrals are all finite over $(-\infty, \infty)$: $\int [S'(u)]^2 du$ and $\int |S'(u)|^{2+\delta} du$ ($\delta > 0$); (iii) for each i ($1 \leq i \leq r$), $\int |u^i S'(u)| du < \infty$ and

$$\int_{-\infty}^{\infty} u^k S'(u) du = \begin{cases} 0, & \text{if } i < r, \\ \alpha_r (\neq 0), & \text{if } i = r, \end{cases}$$

The conditions (C1) and (C2) assume that the conditional density function $f(x|Y=j)$ and the criterion function $\vartheta(\cdot)$ are sufficiently smooth. This requirement is met when X given Y follows a common continuous distribution, such as Normal or Gamma distribution, and the selection of $\vartheta(\cdot)$ corresponds to a relationship measure, such as weighted Kappa, Kendall's τ_b , correct classification rate or maximum absolute determinant (MADET). Conditions (C1) and (C2) can also be satisfied when the distribution of X given Y varies with Y , or given other choices of $\vartheta(\cdot)$. The condition (C3) indicates the technical constraints for the bandwidth h_n . Condition (C4) mimics the usual assumptions for nonparametric kernel smoothing. We can show that the two examples of the smooth function $S(\cdot)$ presented in Section 2 satisfy condition (C4).

In Theorem 1, we show that the non-smooth criterion function $\vartheta(P_n(\vec{d}))$ and its smooth approximation, $\vartheta(\tilde{P}_n(\vec{d}))$, are asymptotically equivalent and the proposed estimator $\widehat{\vec{d}}_s$ is consistent. In Theorem 2, we establish the asymptotic normality of $\widehat{\vec{d}}_s$.

Theorem 1.—Assume conditions (C1)–(C4) hold.

- (i) The smooth criterion function $\vartheta(\tilde{P}_n(\vec{d}))$ and the non-smooth criterion function $\vartheta(P_n(\vec{d}))$ are asymptotically equivalent, i.e.

$$\sup_{\vec{d} \in \Theta} \left| \vartheta \left(\tilde{P}_n(\vec{d}) \right) - \vartheta \left(P_n(\vec{d}) \right) \right| \xrightarrow{a.s.} 0, \quad (7)$$

where " $\xrightarrow{a.s.}$ " denotes almost sure convergence;

- (ii) The nonparametric estimator $\widehat{\vec{d}}_s$ is consistent, i.e. $\widehat{\vec{d}}_s \xrightarrow{P} \vec{d}_0$, where " $\xrightarrow{P}$ " denotes the convergence in probability.

Theorem 2.—Assume that the conditions (C1)–(C4) hold.

- (i) If $nh_n^{2r+1} \rightarrow \infty$ as $n \rightarrow \infty$, then

$$h_n^{-r} \left(\widehat{\vec{d}}_s - \vec{d}_0 \right) \xrightarrow{P} -Q^{-1}(\vec{d}_0) A(\vec{d}_0),$$

where

$$A(\vec{d}_0) = \frac{\alpha_r}{r!} \begin{pmatrix} \sum_{j=1}^L \left(\vartheta'_j(P(\vec{d}_0)) - \vartheta'_{L+j}(P(\vec{d}_0)) \right) f^{(r)}(d_{1,0} | Y_1 = j) P(Y_1 = j) \\ \vdots \\ \sum_{j=1}^L \left(\vartheta'_{(L-2)L+j}(P(\vec{d}_0)) - \vartheta'_{(L-1)L+j}(P(\vec{d}_0)) \right) f^{(r)}(d_{L-1,0} | Y_1 = j) P(Y_1 = j) \end{pmatrix},$$

and $d_{j,0}$ ($j = 1, 2, \dots, L-1$) denote the j th element of $\vec{d}_0$;

- (ii) If nh_n^{2r+3} converges to 0 as $n \rightarrow \infty$, then

$$\sqrt{nh_n} \left[\widehat{\vec{d}}_s - \vec{d}_0 + h_n^r Q^{-1}(\vec{d}_0) A(\vec{d}_0) + o_p(h_n^r) \right] \xrightarrow{d} N \left(0, Q^{-1}(\vec{d}_0) \Pi(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right),$$

where " $\xrightarrow{d}$ " denotes convergence in distribution, and $\Pi(\vec{d}_0)$ is defined in the Appendix. In particular, if nh_n^{2r+1} converges to a finite number λ as $n \rightarrow \infty$, then

$$\sqrt{nh_n} \left(\widehat{\vec{d}}_s - \vec{d}_0 \right) \xrightarrow{d} N \left(-\lambda^{1/2} Q^{-1}(\vec{d}_0) A(\vec{d}_0), Q^{-1}(\vec{d}_0) \Pi(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right).$$

By Theorem 2, the asymptotic bias of $\widehat{\vec{d}}_s$ is of the order $O(h_n^r)$, and the asymptotic variance of $\widehat{\vec{d}}_s$ is of the order $O\left(\frac{1}{nh_n}\right)$. This implies that increasing the bandwidth h_n will increase the bias and reduce the variance; decreasing h_n will decrease the bias and increase the variance. Interestingly, it is also suggested that using a higher-order kernel function will help reduce the bias. Similar findings have been seen in other nonparametric estimation context.

Based on Theorem 2, the asymptotic mean squared error (MSE) of $\widehat{\vec{d}}_s$ is equal to

$$\begin{aligned} \text{AMSE}(\widehat{\vec{d}}_s) &= E(\widehat{\vec{d}}_s - \vec{d}_0)^T (\widehat{\vec{d}}_s - \vec{d}_0) = \text{tr} \left[E(\widehat{\vec{d}}_s - \vec{d}_0) (\widehat{\vec{d}}_s - \vec{d}_0)^T \right] \\ &= h^{2r} A^T(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T Q^{-1}(\vec{d}_0) A(\vec{d}_0) + \frac{1}{nh_n} \text{tr} \left[Q^{-1}(\vec{d}_0) \pi(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right], \end{aligned}$$

where $\text{tr}(\mathbf{B})$ denotes the trace of the matrix $\mathbf{B}$. Taking the derivative with respect to h_n it follows that

$$\frac{d\text{AMSE}(\widehat{\vec{d}}_s)}{dh_n} = 2rh_n^{2r-1} A^T(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T Q^{-1}(\vec{d}_0) A(\vec{d}_0) - \frac{1}{nh_n^2} \text{tr} \left[Q^{-1}(\vec{d}_0) \pi(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right]$$

The solution to $\frac{d\text{AMSE}(\widehat{\vec{d}}_s)}{dh_n} = 0$ (as an equation for h_n) has a closed form,

$$h_n^{\text{opt}} = \left\{ \text{tr} \left[Q^{-1}(\vec{d}_0) \pi(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right] / 2r \text{tr} \left[Q^{-1}(\vec{d}_0) A(\vec{d}_0) A^T(\vec{d}_0) (Q^{-1}(\vec{d}_0))^T \right] \right\}^{1/(2r+1)} n^{-1/(2r+1)}.$$

This result suggests that the optimal convergence rate that $\widehat{\vec{d}}_s$ can achieve is $n^{r/(2r+1)}$. A faster convergence rate would be achieved with the use of a higher-order local distribution function.

3.2. Covariance estimation

Let $\hat{d}_{j,s}$ ($j = 1, 2, \dots, L-1$) denotes the j th element of $\widehat{\vec{d}}_s$. By the results of Theorem 2, the asymptotic covariance matrix of $\widehat{\vec{d}}_s$ can be estimated by

$$\frac{1}{nh_n} \widehat{Q}^{-1}(\widehat{\vec{d}}_s) \widehat{\Pi}(\widehat{\vec{d}}_s) \left(\widehat{Q}^{-1}(\widehat{\vec{d}}_s) \right)^T,$$

where

$$\hat{\Pi}(\widehat{\mathbf{d}}_s) = \frac{1}{nh} \sum_{i=1}^n B_i(\widehat{\mathbf{d}}_s) B_i^T(\widehat{\mathbf{d}}_s),$$

and

$$\begin{aligned} \hat{\mathbf{Q}}(\widehat{\mathbf{d}}_s) &= \hat{\mathbf{H}}(\widehat{\mathbf{d}}_s) \cdot \mathbf{I} \cdot \mathbf{C}(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s)) \cdot \mathbf{I}^T \cdot \hat{\mathbf{H}}^T(\widehat{\mathbf{d}}_s) + \text{diag} \left\{ \sum_{j=1}^L \left(\vartheta'_j(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s)) - \vartheta'_{L+j}(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s)) \right) \right\} \\ &\times \frac{1}{nh_n^2} \sum_{k=1}^n S'' \left(\frac{\hat{d}_{1,s} - X_k}{h_n} \right) I(Y_k = j), \dots, \sum_{j=1}^L \left(\vartheta'_j(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s)) - \vartheta'_{L+j}(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s)) \right) \\ &\times \frac{1}{nh_n^2} \sum_{k=1}^n S'' \left(\frac{\hat{d}_{L-1,s} - X_k}{h_n} \right) I(Y_k = j), \end{aligned}$$

where $\text{diag}\{a_1, \dots, a_n\}$ denotes a diagonal matrix whose diagonal entries starting in the upper left corner are $a_1, \dots, a_n$, and the matrices $\mathbf{I}$, $\mathbf{C}(\tilde{\mathbf{P}}_n(\widehat{\mathbf{d}}_s))$, $\hat{\mathbf{H}}(\widehat{\mathbf{d}}_s)$ and $\mathbf{B}_i^T(\widehat{\mathbf{d}}_s)$ are defined in the Appendix.

3.3. Selection of Local Distribution Function $\mathcal{S}(\cdot)$ and Bandwidth h_n

As shown in Theorem 2, the asymptotic bias of the proposed estimator is related to r and thus the specification of the local distribution function $\mathcal{S}(\cdot)$. The local distribution function $\mathcal{S}(\cdot)$ is also subject to the technical constraints stated in condition (C4). In the following, we present two examples of $\mathcal{S}(\cdot)$ that satisfy condition (C4) with $r=2$ or 4:

- For $r=2$, we may choose $\mathcal{S}(u)$ as the integral of second order Epanechnikov kernel [38], i.e.

$$S(u) = \begin{cases} 1, & \text{if } u > \sqrt{5}, \\ \frac{3}{4\sqrt{5}} \left(u - \frac{1}{15}u^3 + \frac{2\sqrt{5}}{3} \right), & \text{if } -\sqrt{5} \leq u \leq \sqrt{5}, \\ 0, & \text{if } u < -\sqrt{5}. \end{cases} \quad (8)$$

- For $r=4$, we may select $\mathcal{S}(u)$ as

$$S(u) = \begin{cases} 1, & \text{if } u > \sqrt{5}, \\ \frac{3}{4\sqrt{5}} \left(\frac{15}{8}u - \frac{7}{12}u^3 + \frac{7}{200}u^5 + \frac{2}{3}\sqrt{5} \right), & \text{if } -\sqrt{5} \leq u \leq \sqrt{5}, \\ 0, & \text{if } u < -\sqrt{5}, \end{cases} \quad (9)$$

which is the integral of fourth order Epanechnikov kernel [38]. Other types of kernel functions, such as Gaussian kernel, can be used to construct $\mathcal{S}(\cdot)$ satisfying (C4). Refer to [38] for other examples of the local distribution function.

Motivated by Bowman et al. [39], we propose a fully automatic and “data-derived” method to select the bandwidth h_n . Specifically, we select h_n as the minimizer of the following leave-one-out cross-validation function:

$$CV(h_n) = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^L \int \left\{ I(X_k \leq x, Y_k = l) - \frac{1}{n-1} \sum_{k_1 \neq k}^n S\left(\frac{x - X_{k_1}}{h_n}\right) I(Y_{k_1} = l) \right\}^2 dx. \quad (10)$$

Note that the term $(n-1)^{-1} \sum_{k_1 \neq k}^n S\left(\frac{x - X_{k_1}}{h_n}\right) I(Y_{k_1} = l)$ in (10) is a leave-one-out kernel

estimator of $P(X \leq x, Y = l)$ ($l = 1, 2, \dots, L$). Therefore, $CV(h_n)$ evaluates h_n by gauging the resulting discrepancy between the empirical distribution estimator and the leave-one-out kernel estimator of $P(X \leq x, Y = l)$.

When the sample size is large, implementing the above leave-one-out cross validation method can be time consuming. To save computation time, we can alternatively adopt the strategy of K -fold cross validation. That is, we first randomly partition the data into K equally sized subsets, say $S_1, \dots, S_K$. Then we select the bandwidth h_n by minimizing the following cross-validation function:

$$CV^K(h_n) = \frac{1}{K} \sum_{k=1}^K \sum_{i \in S_k} \sum_{l=1}^L \int \left\{ I(X_i \leq x, Y_i = l) - \frac{K}{n(K-1)} \sum_{k_1 \notin S_k}^n S\left(\frac{x - X_{k_1}}{h_n}\right) I(Y_{k_1} = l) \right\}^2 dx. \quad (11)$$

In Section 4, we use the leave-one-out method for the simulation studies and the K -fold cross-validation for the real data example where the sample size is large (i.e. $n = 1,430$).

4. Numerical Studies

4.1. Simulation studies

We conduct extensive simulations to evaluate the finite-sample performance of the proposed method. We consider weighted kappa coefficient (wkappa), Kendall's τ_b based measure method (kendall), correct classification (CC) and maximum absolute determinant (MADET) as the criterion function for defining the optimal cut-points, which are respectively denoted by $\vartheta_K(\cdot)$, $\vartheta_{CC}(\cdot)$, $\vartheta_\tau(\cdot)$ and $\vartheta_{MD}(\cdot)$. Specifically,

$$\vartheta_K(P(\vec{d})) = \frac{\sum_{i=1}^L \sum_{j=1}^L w_{ij} p_{ij}(\vec{d}) - \sum_{i=1}^L \sum_{j=1}^L w_{ij} p_{i \cdot}(\vec{d}) p_{\cdot j}(\vec{d})}{1 - \sum_{i=1}^L \sum_{j=1}^L w_{ij} p_{i \cdot}(\vec{d}) p_{\cdot j}(\vec{d})},$$

where $p_{i \cdot}(\vec{d}) = \sum_{j=1}^L p_{ij}(\vec{d})$, $p_{\cdot j}(\vec{d}) = \sum_{i=1}^L p_{ij}(\vec{d})$ and $\{w_{ij}\}_{i,j=1}^L$ are the weight which denote the discrepancy between two categories. In addition, $\vartheta_{CC}(P(\vec{d})) = \sum_{i=1}^L p_{ii}(\vec{d})$

$$\vartheta_{\tau}(P(\vec{d}))$$

$$= \frac{\sum_{i=1}^{L-1} \sum_{j=1}^{L-1} p_{ij}(\vec{d}) \left(\sum_{t=i+1}^L \sum_{m=j+1}^L p_{tm}(\vec{d}) \right) - \sum_{i=2}^{L-1} \sum_{j=2}^L p_{ij}(\vec{d}) \left(\sum_{t=i+1}^L \sum_{m=1}^{j-1} p_{tm}(\vec{d}) \right)}{\frac{1}{2} \left\{ \left(1 - \sum_{i=1}^L p_{i \cdot}(\vec{d})^2 \right) \left(1 - \sum_{j=1}^L p_{\cdot j}(\vec{d})^2 \right) \right\}^{1/2}}$$

and

$$\vartheta_{MD}(P(\vec{d})) = \left| \det(\mathcal{P}(\vec{d})) \right|,$$

where $\det(\mathbf{B})$ denotes the determinant of the matrix $\mathbf{B}$, and $\mathcal{P}(\vec{d})$ is a matrix whose (i, j) -element is equal to $\left\{ p_{\cdot j}(\vec{d}) \right\}^{-1} p_{ij}(\vec{d})$, $i, j = 1, 2, \dots, L$.

We first examine six data scenarios with various combinations of L and the conditional distribution of X given Y . The data scenarios are described as follows.

Scenario 1: With $L = 2$, Y is a binary random variable taking values 3 and 4 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = l \sim N(2l, 0.6^2)$.

Scenario 2: With $L = 2$, Y is a binary random variable taking values 3 and 4 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = 3 \sim \text{Gamma}(2, 1)$ and $X|Y = 4 \sim \text{Gamma}(5, 0.7)$.

Scenario 3: With $L = 3$, Y is a discrete random variable taking values 3, 4 and 5 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = l \sim N(2l, 0.6^2)$.

Scenario 4: With $L = 3$, Y is a discrete random variable taking values 3, 4 and 5 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = 3 \sim \text{Gamma}(2, 1)$, $X|Y = 4 \sim \text{Gamma}(5, 0.7)$ and $X|Y = 5 \sim \text{Gamma}(7, 0.5)$.

Scenario 5: With $L = 4$, Y is a discrete random variable taking values 3, 4, 5 and 6 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = l \sim N(2l, 0.6^2)$.

Scenario 6: With $L = 4$, Y is a discrete random variable taking values 3, 4, 5 and 6 with equal probabilities. The conditional distribution of X given Y is specified as $X|Y = 3 \sim$

$\text{Gamma}(2, 1)$, $X|Y=4 \sim \text{Gamma}(5, 0.7)$, $X|Y=5 \sim \text{Gamma}(7, 0.5)$ and $X|Y=6 \sim \text{Gamma}(9, 0.3)$.

Note that, in Scenarios 1, 3 and 5, X given Y is normally distributed with the same variance but different means. Under these scenarios, the optimal set of cut-points is (7, 8, 9) under all considered criterion functions. In Scenarios 2, 4 and 6, the conditional distributions of X given Y are not symmetric, sometimes are highly skewed, and have means and variances changing with the value of Y . The optimal sets of cut-points in these settings do not have simple analytical form and vary with the criterion function. We assess the true optimal cut-points via Monte-Carlo approximation given the knowledge of the underlying distributions of (X, Y) . For each data scenario, we consider sample sizes $n = 100, 200$ and 300 with 1000 replicates.

We first implement the proposed method with $\mathcal{S}(\cdot)$ set as the integral of second order Epanechnikov kernel ($r=2$). We select bandwidth h_n based on the leave-one-out cross-validation method presented in Section 3. Table 1 presents the results for Scenarios 1–2, including empirical bias (BIAS), the empirical relative bias (RBIAS), empirical standard deviation (SE), average estimated standard deviation (ASE), average root mean square error (RMSE), empirical coverage probability of 95% confidence interval (CP) and computation time in hour for 1000 simulation runs (CT). Here, the relative bias is defined as the ratio of bias to true estimand. “Smooth” and “Non-smooth” refer to, respectively, the simulation results obtained by the proposed smooth nonparametric approach and Peng et al. [25]’s approach.

From Table 1, we observe that for Scenario 1, the proposed approach yields very small empirical biases. The SE’s and the corresponding ASE’s based on all criterion functions agree quite well for all sample sizes. The CP’s are close to the nominal value 95%. The observations for Scenario 2 are similar except those on BIAS and CP. The empirical biases are quite large when $n = 100$, while the empirical relative biases are still under control (e.g. all are below 0.05). When the sample size increases, the empirical biases become smaller and the coverage probabilities of 95% confidence intervals get closer to 0.95. Comparing the proposed method with Peng et al. [25]’s method, we find that the empirical biases associated with Peng et al. [25]’s non-smooth method are consistently small for both Scenario 1 and Scenario 2. On the other hand, we note that the empirical standard deviations of the new cut-point estimates are nearly half of those from the non-smooth approach. Moreover, we observe 25–50% reduction in RMSE (which accounts for both estimation bias and variability) from using the proposed smooth method. These results confirm the efficiency gains associated with the proposed smooth approach, as indicated by Theorem 2. They also illustrate that with a smaller sample size, the tradeoff between bias and variance resulted from kernel smoothing can be more evident. In terms of computational speed, the non-smooth approach is much faster than the smooth approach. This is because the non-smooth approach only involves the maximization of a univariate objective function when $L = 2$; grid-searching its maximum can be done efficiently by bisection or other algorithm. At the same time, the smooth approach involves repeated calculations for bandwidth selection and hence takes longer time.

The simulation results for Scenario 3 with $L = 3$ and normally distributed X are presented in Table 2 in the same manner as in Table 1. In this data scenario, we observe very good finite-sample performance of the proposed method in terms of small empirical biases, a good agreement between standard deviation estimates and empirical standard deviations and empirical coverage probabilities close to 95%. Again we observe that using the new smooth approach can result in up to 50% smaller RMSE compared to Peng et al. [25]'s approach. Moreover, the smooth approach is computationally much faster than the non-smooth approach, despite the smooth approach involves cross-validation to select h_n . When the criterion function corresponds to Kendall's τ_b , MADET or weighted Kappa, the ratios of CT of the smooth approach to that of the non-smooth approach range from 0.07 to 0.32. When misclassification rate is adopted as the criterion function, the computational gains are smaller. The saving in computation time becomes more substantial when the sample size increases. These results clearly demonstrate the advantages of the smooth approach over the non-smooth approach in both statistical and computational efficiency.

Table 3 presents the simulation results for Scenario 4, where $L = 3$ and the distribution of X given Y can be skewed for some values of Y and has mean and variance changing with Y . In this case, the proposed smooth approach yields larger biases than those from the non-smooth approach, while remaining to produce smaller RMSE's; the ratio of RMSE from the smooth approach to RMSE from the non-smooth approach ranges from 45–83%. Like in Scenario 3, we consistently observe considerable computational gains associated with the smooth approach in Scenario 4.

Tables 4 and 5 present the simulation results for Scenarios 5 and 6 respectively. In these settings with $L = 4$, it is too time-consuming to obtain the results from implementing Peng et al. [25]'s non-smooth procedure with $n = 100, 200$ and 300 (e.g. final results not being obtained after one week). Therefore, for these data settings, we just present the simulation results for the proposed smooth method and omit its comparisons with the non-smooth method. As shown by Tables 4 and 5, the computation time for 1000 simulation runs based on the smooth approach are very reasonable, suggesting the much improved computational feasibility of the proposed smooth method in settings with relatively large sample sizes and more than three ordinal categories. The observations on the estimation performance in general are similar to those for the cases with $L = 3$. The empirical bias and coverage probabilities are better in data settings where X given Y are normally distributed than the data settings where X given Y has asymmetric and possibly skewed distributions. In the latter case, the empirical relative bias is mostly under 0.05, and the RMSE has reasonable magnitude and decreases with the sample size. These endorse the utility of the proposed method in the realistic situations depicted by Scenarios 5 and 6.

We consider two additional data scenarios where Y does not have equal probabilities across its categories:

Scenario 7: With $L = 3$, Y is a discrete random variable taking values 3, 4 and 5 with probabilities 0.4, 0.1 and 0.5 respectively. The conditional distribution of X given Y is specified as $X|Y = I \sim N(2I, 0.6^2)$.

Scenario 8: With $L = 3$, Y is a discrete random variable taking values 3, 4 and 5 with probabilities 0.4, 0.1 and 0.5 respectively. The conditional distribution of X given Y is specified as $X|Y = 3 \sim \text{Gamma}(2, 1)$, $X|Y = 4 \sim \text{Gamma}(5, 0.7)$ and $X|Y = 5 \sim \text{Gamma}(7, 0.5)$.

Table 6 present the simulation results for Scenarios 7 and 8 with $n = 300$. Comparing these results with their counterparts in Tables 2 and 3, we note that the performances in terms of bias and computation time are quite comparable, while the SE's are generally larger when Y has unequal probabilities across its categories. The larger estimation variability in Scenario 7 or 8 may be explained by the smaller sample size in the category with $Y = 4$, which results in the reduced accuracy of identifying the cut-points that differentiate the group with $Y = 4$ from the groups with $Y = 3$ and $Y = 5$.

To assess how the specification of $\mathcal{S}(\cdot)$ influences the performance of the proposed method, we conduct simulations for Scenarios 3 and 4 with $\mathcal{S}(\cdot)$ being the integral of fourth order Epanechnikov kernel ($r = 4$), the integral of second order Gaussian kernel ($r = 2$) or the integral of fourth order Gaussian kernel ($r = 4$). The corresponding results (given $n = 100$ and 200) for Scenario 3 and Scenario 4 are respectively summarized in Table 7 and Table 8. Comparing Tables 7–8 with Tables 2–3, we find that RMSE's vary slightly by the type of kernel (i.e. Epanechnikov kernel vs. Gaussian kernel). With $n = 200$, RMSE's show a decreasing trend as the order r increases. This conforms to our theoretical results (see Theorem 2), which suggest a faster convergence rate of $\widehat{d}$ with a larger r . In addition, we note that using a higher order kernel generally incurs longer computation time.

As suggested by our simulation results, the performance of the proposed method may vary across different data scenarios and different choices of criterion functions and local distribution functions. This phenomenon is consistent with our theoretical results, which show a clear dependency of the asymptotic bias and variance of $\widehat{d}_s$ with the choice of $\mathcal{K}(\cdot)$ and $\mathcal{S}(\cdot)$ and the distribution of X given Y . Our simulation studies demonstrate reasonably good finite-sample performance of the proposed method in various data settings, thereby supporting its general applications to real data sets.

4.2. Application to the Diabetes and Depression Study

Diabetes and Depression study [40] was conducted in 1,430 African Americans with type II diabetes patients. The aim of this study was to determine the prevalence of depression in this population and its association with socioeconomic determinants, adherence to treatment and glycemic control. Both the Mini International Neuropsychiatric Interview (MINI) diagnostic interview [41] and the Zung self-rating depression scale [2] were administered in this study. MINI is a widely used psychiatric structured diagnostic interview instruments. This approach in general provide an accurate diagnose and has well-developed graded severity guidelines, including no depression, mild depression and markedly severe depression [41]. However, it is designed for licensed professionals or well-trained interviewers and thus is time-consuming and costly. In contrast, Zung is a survey with 20 items that rates the affective, psychological and somatic symptoms associated with depression. Unlike MINI,

Zung does not need a high level of literacy and can be finished in 5–10 minutes. However, there still lacks a routine, reliable method for determining the cut-points in the Zung score to reflect graded severity of depression. The application of the proposed method will help address this limitation.

In this dataset, the average and median Zung scores are 45.4 and 43.8 respectively, and the inter-quartile range for Zung is (35.0, 55.0). We apply the proposed method to classify the Zung scale into three categories: no depression, mild depression and moderate to severe depression. In our analysis, we choose the local distribution function as the integral of second order Epanechnikov kernel function, given by (8). We use the 10-fold cross-validation method to select bandwidth, and the selected bandwidth is $h_n = 1.428$. In Table 9, we present the estimation results from Peng et al. [25]’s method (labeled by “Non-smooth”) and the proposed method (labeled by “Smooth”), based on the criterion functions, Kendall’s τ_b based measure method (ϑ_τ), maximum absolute determinant (MADET, ϑ_{MD}), weighted kappa coefficient (ϑ_κ) and correct classification rate (ϑ_{CC}).

The analysis results are presented in Table 9, where EST, SE, CI and CT_s denote, respectively, the abbreviations for estimated optimal cut-point, standard error, 95% confidence interval and computational time (in second). From Table 9, we first note that the non-smooth approach and the smooth approach produce quite similar cut-point estimates under each given criterion function. The estimated standard errors (SE) from the smooth approach are generally smaller than those from the non-smooth approach.

Regarding the first cut-point that differentiates no depression versus mild depression, the estimates based on the criteria, weighted Kappa, Kendall’s τ_b and correct classification rate, are very consistent, all around 56. For the second cut-point, adopting Kendall’s τ_b yields considerably larger estimates and standard errors as compared to using weighted Kappa and correct classification rate (e.g. by the non-smooth approach, the estimates are 63.1 vs. 57 and 57, SE’s are 4.4 vs. 1.4 and 1.2). In Figure 1, we plot the empirical criterion function $\hat{\vartheta}_\tau(P_n(\vec{d}))$. The plot suggests a bi-modal underlying criterion function $\vartheta_\tau(P_n(\vec{d}))$, while the identifiability of the optimal cut-points requires a unimodal $\vartheta_\tau(\cdot)$. By this observation, we would caution the use of the cut-point estimates derived based on Kendall’s τ_b in this data example.

From Table 9, we also note that estimates based on the MADET criterion are quite different from those based on the other criterion functions with respect to both cut-points. The plot of the empirical criterion function $\hat{\vartheta}_{MD}(P_n(\vec{d}))$ (see Figure 1) nevertheless demonstrates a unimodal shape. By this observation, we conjecture that the observed discrepancies in the cut-point estimates may simply reflect that the optimal cut-points defined based on MADET are different from those defined based on weighted Kappa or correct classification rate. This may be because the relationship targeted by MADET somewhat deviates from the association and agreement respectively targeted by weighted Kappa and correct classification rate.

In view of the low prevalence rate of mild depression, we combine the mild depression group with the no depression group and conduct analyses to identify the cut-point to differentiate no or mild depression versus moderate or severe depression. The results are presented in Table 10. The cut-point estimates obtained from the proposed smooth method are only slightly different from those from Peng et al. [25]'s non-smooth approach. The estimated optimal cut-points based on the criteria, weighted Kappa, Kendall's τ_b and correct classification rate, are all close to 57, while the estimated cut-point based on MADET is much lower, around 48. In addition, confidence intervals produced by the smooth approach are generally narrower than those from the non-smooth approach. The CT_s 's are larger for the smooth approach because it involves efforts for bandwidth selection. This observation is consistent with our Monte-Carlo simulation results for cases with $L = 2$.

5. Concluding Remarks

The problem of determining cut-points of a continuous scale according to an establish categorical scale is often encountered in practice. This work adopts Peng et al. [25]'s perspective to define optimal cut-points based on a pre-specified criterion. We propose new estimation and inference procedure for the optimal cut-points, which have substantially improved computational feasibility and asymptotic efficiency over the existing method.

It is worth emphasizing that under Peng et al. [25]'s framework for determining cut-points, the interpretation of the proposed cut-point estimates should be aligned to the adopted criterion function that, for example, captures the agreement or association between the categorized continuous scale and the established ordinal scale. The criterion function should be selected according to the scientific problem at hand and/or the data context. Deciding the most suitable criterion function for identifying cut-points based on a given dataset is an interesting problem but is beyond out of the scope of this work.

As noted from the simulation studies, the numerical performance of the proposed estimator varies by the choice of criterion function $\mathcal{H}(\cdot)$ under the same data set-up. Such a phenomenon is consistent with our asymptotic results in Theorems 1–2, which indicate the asymptotic bias and variance of $\widehat{\vec{d}}_s$ depend on the shape of $\mathcal{H}(\cdot)$; see the definitions of $Q(\vec{d}_0)$ and $A(\vec{d}_0)$. A similar observation was reported by Peng et al. [25], which investigated the binary case with $L = 2$, and revealed that $\mathcal{H}(P(\vec{d}))$, which is flatter around $\vec{d}_0$, tends to yield larger variability in estimating $\vec{d}_0$ (see Fig. 1 of Peng et al. [25]). It is worth emphasizing that the interpretation of $\vec{d}_0$ is contingent upon the choice of $\mathcal{H}(\cdot)$. That is, $\vec{d}_0$ represents the set of cut-points that maximize a relationship measure between Y and categorized X determined by $\mathcal{H}(\cdot)$ (e.g. agreement, association, accuracy). Therefore, one needs to exercise caution when comparing $\widehat{\vec{d}}_s$ and its associated variability across different choices of $\mathcal{H}(\cdot)$.

Acknowledgements

This research project was supported by grants from National Institute of Health (R01MH079448, R01HL113548 and R01MH105561). Dr. Qiu's work was also supported by the Education and Scientific Research Projects of Young and Middle-aged Teachers in Fujian Province, China (Grant No. JAT160027), the Natural Science Foundation of Fujian Province, China (Grant No. 2017J01002).

Appendix A:: Proofs of Theorems

We first introduce necessary notation and present the specific form of $Q(\vec{d})$. Given two matrices $A = (a_{ij})_{m \times n} = [a_1, \dots, a_n] \in \mathbb{R}^{m \times n}$ and $B = [b_1, \dots, b_p] \in \mathbb{R}^{p \times q}$, write the Kronecker product of A and B as $A \otimes B$, which is equal to

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix},$$

and has dimension, $mp \times nq$. When the matrices A and B have the same number of columns (i.e. $n = q$), the Khatri-Rao product [42] is defined as $A \odot B = [a_1 \otimes b_1, a_2 \otimes b_2, \dots, a_n \otimes b_n]$, which is the mp -by- n columnwise Kronecker product A and B . For $\forall \vec{d} \in \Theta$, define

$$C(P(\vec{d})) = \begin{pmatrix} C_{11}(P(\vec{d})) & C_{12}(P(\vec{d})) & \cdots & C_{1L}(P(\vec{d})) \\ C_{21}(P(\vec{d})) & C_{22}(P(\vec{d})) & \cdots & C_{2L}(P(\vec{d})) \\ \vdots & \vdots & \ddots & \vdots \\ C_{L1}(P(\vec{d})) & C_{L2}(P(\vec{d})) & \cdots & C_{LL}(P(\vec{d})) \end{pmatrix},$$

where for $i, j = 1, 2, \dots, L$,

$$C_{ij}(P(\vec{d})) = \begin{pmatrix} \vartheta''_{((i-1)L+1)((j-1)L+1)}(P(\vec{d})) & \cdots & \vartheta''_{((i-1)L+1)(jL)}(P(\vec{d})) \\ \vdots & & \vdots \\ \vartheta''_{(iL)((j-1)L+1)}(P(\vec{d})) & \cdots & \vartheta''_{(iL)(jL)}(P(\vec{d})) \end{pmatrix}.$$

Let I_L denotes the $L \times L$ identity matrix and a $(L^2 - L) \times L^2$ matrix

$$I = \begin{pmatrix} I_L & -I_L & 0 & \cdots & 0 & 0 \\ 0 & I_L & -I_L & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_L & -I_L \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \otimes I_L.$$

For $i = 1, 2, \dots, L-1$, let

$$H_i = \begin{pmatrix} F'(d_i|Y=1)P(Y=1) \\ \vdots \\ F'(d_i|Y=1)P(Y=L) \end{pmatrix},$$

and for any $\vec{d} \in \Theta$, define

$$H(\vec{d}) = \begin{pmatrix} H_1^T & 0 & \cdots & 0 \\ 0 & H_2^T & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & H_{L-1}^T \end{pmatrix} = (I_{L-1} \odot (H_1, \dots, H_{L-1}))^T,$$

and

$$\begin{aligned} Q(\vec{d}) = & H(\vec{d}) \cdot I \cdot C(P(\vec{d})) \cdot I^T \cdot H^T(\vec{d}) + \text{diag} \left\{ \sum_{j=1}^L (\vartheta'_j(P(\vec{d})) - \vartheta'_{L+j}(P(\vec{d}))) \right. \\ & \times F''(d_1|Y=j)P(Y=j), \dots, \sum_{j=1}^L (\vartheta'_{(L-2)L+j}(P(\vec{d})) - \vartheta'_{(L-1)L+j}(P(\vec{d}))) \\ & \left. \times F''(d_{L-1}|Y=j)P(Y=j) \right\}, \end{aligned}$$

where $d_j (j=1, 2, \dots, L-1)$ denotes the j th element of $\vec{d}$, $F'(x|Y=j)$ and $F''(x|Y=j)$ are, respectively, the first-order derivative and second-order derivative of the conditional distribution function of X given $Y=j$, $F(x|Y=j)$.

Next we present the proofs of the Theorem 1 and Theorem 2.

Proof of Theorem 1: (i) Recall that $P_n(\vec{d}) = (\hat{p}_{11}(\vec{d}), \dots, \hat{p}_{1L}(\vec{d}), \dots, \hat{p}_{L1}(\vec{d}), \dots, \hat{p}_{LL}(\vec{d}))^T$, $\tilde{P}_n(\vec{d}) = (\tilde{p}_{11}(\vec{d}), \dots, \tilde{p}_{1L}(\vec{d}), \dots, \tilde{p}_{L1}(\vec{d}), \dots, \tilde{p}_{LL}(\vec{d}))^T$. Under the condition (C1), in order to prove (i) in Theorem 1, it suffices to prove that for $i, j=1, 2, \dots, L$,

$$\sup_{\vec{d} \in \Theta} |\hat{p}_{ij}(\vec{d}) - \tilde{p}_{ij}(\vec{d})| \xrightarrow{a.s.} 0. \quad (12)$$

By the definitions of $\tilde{p}_{ij}(\vec{d})$ and $\hat{p}_{ij}(\vec{d})$, we have

$$\begin{aligned} \left| \hat{p}_{ij}(\vec{d}) - \tilde{p}_{ij}(\vec{d}) \right| &\leq \frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_i - X_k}{h_n}\right) - I(X_k \leq d_i) \right| + \\ &\quad \frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_{i-1} - X_k}{h_n}\right) - I(X_k \leq d_{i-1}) \right|. \end{aligned} \quad (13)$$

Thus we only need to prove that for $i = 1, 2, \dots, L$,

$$\sup_{d_i \in [x_L, x_U]} \frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_i - X_k}{h_n}\right) - I(X_k \leq d_i) \right| \xrightarrow{a.s.} 0. \quad (14)$$

First, we note that for any $\eta > 0$,

$$\begin{aligned} &\frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_i - X_k}{h_n}\right) - I(X_k \leq d_i) \right| \\ &= \frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_i - X_k}{h_n}\right) - I(X_k \leq d_i) \right| I(|d_i - X_k| \geq \eta) \\ &\quad + \frac{1}{n} \sum_{k=1}^n \left| S\left(\frac{d_i - X_k}{h_n}\right) - I(X_k \leq d_i) \right| I(|d_i - X_k| < \eta) \\ &\equiv I_1 + I_2. \end{aligned}$$

For I_1 , since the real-valued function classes, $F_1 = \{S((d_i - x)/h_n) : d_i \in [x_L, x_U]\}$, $F_2 = \{I(x \leq d_i) : d_i \in [x_L, x_U]\}$ and $F_3 = \{I(|d_i - x| \geq \eta) : d_i \in [x_L, x_U]\}$, are Euclidean with constant envelope function $F \equiv 1$, the function class $F = \{ |S((d_i - x)/h_n) - I(X_k \leq d_i)| I(|d_i - X_k| \geq \eta) : d_i \in [x_L, x_U] \}$ is also Euclidean with envelope function $F \equiv 1$ [43]. Thus, by conditions (C3) and (C4) and the Lebesgue's dominated theorem, I_1 converges to 0 uniformly over $[x_L, x_U]$ for any $\eta > 0$.

For I_2 , we have

$$I_2 \leq \frac{1}{n} \sum_{k=1}^n I(|d_i - X_k| < \eta). \quad (15)$$

By the Glivenko-Cantelli theorem [44], the right-hand side of (15) converges almost surely to $P(|d_i - X| < \eta)$ over $[x_L, x_U]$. Letting $\eta \rightarrow 0$, I_2 converges to 0 under (C2). This completes the proof of (i) in Theorem 1. We can prove (ii) using the same arguments of Peng et al. [25].

Lemma 1.

If the conditions (C1)–(C4) hold. Then

i.

$$\lim_{n \rightarrow \infty} E \left[h_n^{-r} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(\mathbf{P}(\vec{d})) \right] \Big|_{\vec{d} = \vec{d}_0} = A(\vec{d}_0).$$

ii.

$$\lim_{n \rightarrow \infty} Var \left[(nh_n)^{1/2} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(\mathbf{P}(\vec{d})) \right] \Big|_{\vec{d} = \vec{d}_0} = \Pi(\vec{d}_0) = (\Pi_{lm}(\vec{d}_0)) \in \mathbb{R}^{(L-1) \times (L-1)},$$

where

$$\Pi_{ll}(\vec{d}_0) = \sum_{j=1}^L \left(\vartheta'_{((l-1)L+j)}(\mathbf{P}(\vec{d}_0)) - \vartheta'_{lL+j}(\mathbf{P}(\vec{d}_0)) \right)^2 f(d_{l,0}|Y_1=j)P(Y_1=j) \text{ and } \int S'^2(v)dv,$$

for $l = m$ ($l, m = 1, 2, \dots, L-1$),

$$\begin{aligned} \Pi_{lm}(\vec{d}_0) &= \lim_{n \rightarrow \infty} \frac{1}{h_n} \sum_{j=1}^L \left(\vartheta'_{((l-1)L+j)}(\mathbf{P}(\vec{d}_0)) - \vartheta'_{lL+j}(\mathbf{P}(\vec{d}_0)) \right) \\ &\times \left(\vartheta'_{(m-1)L+j}(\mathbf{P}(\vec{d}_0)) - \vartheta'_{mL+j}(\mathbf{P}(\vec{d}_0)) \right) E \left[S' \left(\frac{d_{l,0} - X_1}{h_n} \right) S' \left(\frac{d_{m,0} - X_1}{h_n} \right) I(Y_1=j) \right]. \end{aligned}$$

Proof of Lemma 1: (i) Let $\vartheta'_k(\cdot)$ ($k = 1, 2, \dots, L^2$) denote the partial derivative of the function $\vartheta(\cdot)$ with respect to k th element, and $\vartheta'(\cdot) = \left(\vartheta'_1(\cdot), \dots, \vartheta'_L(\cdot), \dots, \vartheta'_{L(L-1)+1}(\cdot), \dots, \vartheta'_{L^2}(\cdot) \right)^T$.

Moreover, for $i, j = 1, 2, \dots, L$ and $k = 1, 2, \dots, n$, define

$$\tilde{P}_{ijk}(\vec{d}) = \left[S \left(\frac{d_i - X_k}{h_n} \right) - S \left(\frac{d_{i-1} - X_k}{h_n} \right) \right] I(Y_k = j).$$

Then it follows that

$$\begin{aligned}
& \left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(\mathbf{P}(\vec{d}_0)) = \frac{1}{n} \sum_{k=1}^n \begin{pmatrix} \sum_{i=1}^L \sum_{j=1}^L \vartheta'_{(i-1)L+j}(\mathbf{P}(\vec{d}_0)) \frac{\partial \tilde{\mathbf{P}}_{ijk}(\vec{d})}{\partial d_1} \Big|_{\vec{d} = \vec{d}_0} \\ \vdots \\ \sum_{i=1}^L \sum_{j=1}^L \vartheta'_{(i-1)L+j}(\mathbf{P}(\vec{d}_0)) \frac{\partial \tilde{\mathbf{P}}_{ijk}(\vec{d})}{\partial d_{L-1}} \Big|_{\vec{d} = \vec{d}_0} \end{pmatrix} \\
& = \frac{1}{nh_n} \sum_{k=1}^n \begin{pmatrix} \sum_{j=1}^L \left(\vartheta'_j(\mathbf{P}(\vec{d}_0)) - \vartheta'_{L+j}(\mathbf{P}(\vec{d}_0)) \right) S' \left(\frac{d_{1,0} - X_k}{h_n} \right) I(Y_k = j) \\ \vdots \\ \sum_{j=1}^L \left(\vartheta'_{(L-2)L+j}(\mathbf{P}(\vec{d}_0)) - \vartheta'_{(L-1)L+j}(\mathbf{P}(\vec{d}_0)) \right) S' \left(\frac{d_{L-1,0} - X_k}{h_n} \right) I(Y_k = j) \end{pmatrix}.
\end{aligned}
\tag{16}$$

Therefore,

$$\begin{aligned}
& E \left[\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(\mathbf{P}(\vec{d}_0)) \right] \\
& = \frac{1}{h_n} \begin{pmatrix} \sum_{j=1}^L \left(\vartheta'_j(\mathbf{P}(\vec{d}_0)) - \vartheta'_{L+j}(\mathbf{P}(\vec{d}_0)) \right) E \left[S' \left(\frac{d_{1,0} - X_1}{h_n} \right) I(Y_1 = j) \right] \\ \vdots \\ \sum_{j=1}^L \left(\vartheta'_{(L-2)L+j}(\mathbf{P}(\vec{d}_0)) - \vartheta'_{(L-1)L+j}(\mathbf{P}(\vec{d}_0)) \right) E \left[S' \left(\frac{d_{L-1,0} - X_1}{h_n} \right) I(Y_1 = j) \right] \end{pmatrix}.
\end{aligned}
\tag{17}$$

Using the Taylor expansion, for $l = 1, 2, \dots, L-1, j = 1, 2, \dots, L$, we have

$$\begin{aligned}
 & \frac{1}{h_n} E \left[S' \left(\frac{d_{l,0} - X_1}{h_n} \right) I(Y_1 = j) \right] \\
 &= \frac{1}{h_n} \int S' \left(\frac{d_{l,0} - x_1}{h_n} \right) f(x_1 | Y_1 = j) P(Y_1 = j) dx_1 \\
 &= \int S'(v) f(vh_n + d_{l,0} | Y_1 = j) P(Y_1 = j) dv \\
 &= \int S'(v) \left[f(d_{l,0} | Y_1 = j) + f'(d_{l,0} | Y_1 = j)vh_n + \dots + \frac{1}{(r-1)!} f^{(r-1)}(d_{l,0} | Y_1 = j)v^{r-1}h_n^{r-1} \right. \\
 &\quad \left. + \frac{1}{r!} f^{(r)}(d_{l,0} | Y_1 = j)v^r h_n^r + O(h_n^{r+1}) \right] P(Y_1 = j) dv \\
 &= f(d_{l,0} | Y_1 = j) P(Y_1 = j) + \frac{h_n^r}{r!} f^{(r)}(d_{l,0} | Y_1 = j) P(Y_1 = j) \alpha_r + O(h_n^{r+1}).
 \end{aligned}$$

(18)

Plugging (18) into (17), we obtain

$$\begin{aligned}
 E \left[h_n^{-r} \frac{\partial \tilde{P}_n^T(\vec{d})}{\partial \vec{d}} \right]_{\vec{d} = \vec{d}_0} g'(P(\vec{d}_0)) &= h_n^{-r} \frac{\partial P(\vec{d})}{\partial \vec{d}} \bigg|_{\vec{d} = \vec{d}_0} g'(P(\vec{d}_0)) + A(\vec{d}_0) + O(h_n) = A \\
 &(\vec{d}_0) + o_p(1),
 \end{aligned}$$

(19)

where the last equation follows from that $\vec{d}_0$ is the maximizer of $g(P(\vec{d}))$. We conclude the proof of Lemma 1 (i).

(ii) For $i = 1, 2, \dots, n, j = 1, 2, \dots, L$, define

$$B_i(\vec{d}_0) = \begin{pmatrix} \sum_{j=1}^L \left(g'_j(P(\vec{d}_0)) - g'_{L+j}(P(\vec{d}_0)) \right) S' \left(\frac{d_{1,0} - X_i}{h_n} \right) I(Y_i = j) \\ \vdots \\ \sum_{j=1}^L \left(g'_{(L-2)L+j}(P(\vec{d}_0)) - g'_{(L-1)L+j}(P(\vec{d}_0)) \right) S' \left(\frac{d_{L-1,0} - X_i}{h_n} \right) I(Y_i = j) \end{pmatrix}, \quad (20)$$

and

$$B_{ij}(\vec{d}_0) = \begin{pmatrix} \left(\vartheta'_j(P(\vec{d}_0)) - \vartheta'_{L+j}(P(\vec{d}_0)) \right) S' \left(\frac{d_{1,0} - X_i}{h_n} \right) I(Y_i = j) \\ \vdots \\ \left(\vartheta'_{(L-2)L+j}(P(\vec{d}_0)) - \vartheta'_{(L-1)L+j}(P(\vec{d}_0)) \right) S' \left(\frac{d_{L-1,0} - X_i}{h_n} \right) I(Y_i = j) \end{pmatrix}. \quad (21)$$

Then by (16), it follows that

$$\begin{aligned} & \text{Var} \left[(nh_n)^{1/2} \frac{\partial \tilde{P}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) \Big| \vec{d} = \vec{d}_0 \right] \\ &= \frac{1}{h_n} \text{Var} \left(B_1(\vec{d}_0) \right) = \frac{1}{h_n} E \left[\left(B_1(\vec{d}_0) \right) B_1^T(\vec{d}_0) \right] - \frac{1}{h_n} \left[E \left(B_1(\vec{d}_0) \right) \right] \left[E \left(B_1(\vec{d}_0) \right) \right]^T. \end{aligned} \quad (22)$$

For the second term in right hand side of (22), given (19), by standard manipulations, we have $E(B_1(\vec{d}_0)) = O(h_n^r)$, and thus $\frac{1}{h_n} \left[E(B_1(\vec{d}_0)) \right] \left[E(B_1(\vec{d}_0)) \right]^T = O(h_n^{2r-1}) = o(1)$ For the first term in right hand side of (22),

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{h_n} E \left[B_1(\vec{d}_0) B_1^T(\vec{d}_0) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{h_n} \sum_{j=1}^L E \left[B_{1j}(\vec{d}_0) B_{1j}^T(\vec{d}_0) \right] = (\Pi_{lm}(\vec{d}_0)) \in \mathbb{R}^{(L-1) \times (L-1)}, \end{aligned} \quad (23)$$

where, for $l = 1, 2, \dots, L-1$,

$$\begin{aligned} & \Pi_{ll}(\vec{d}_0) \\ &= \lim_{n \rightarrow \infty} \frac{1}{h_n} \sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(P(\vec{d}_0)) - \vartheta'_{lL+j}(P(\vec{d}_0)) \right)^2 E \left[S' \left(\frac{d_{l,0} - X_1}{h_n} \right) I(Y_1 = j) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{h_n} \sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(P(\vec{d}_0)) - \vartheta'_{lL+j}(P(\vec{d}_0)) \right)^2 \\ & \quad \times \int S' \left(\frac{d_{l,0} - x_1}{h_n} \right) f(x_1 | Y_1 = j) P(Y_1 = j) dx_1 \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(\vec{d}_0) - \vartheta'_{lL+j}(\vec{d}_0) \right)^2 \\
&\quad \times \int S'2(v) f(d_{l,0} + v h_n | Y_1 = j) P(Y_1 = j) dv \\
&= \lim_{n \rightarrow \infty} \left[\sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(\vec{d}_0) - \vartheta'_{lL+j}(\vec{d}_0) \right)^2 \right. \\
&\quad \times f(d_{l,0} | Y_1 = j) P(Y_1 = j) \int S'2(v) dv + O(h_n) \Big] \\
&= \sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(\vec{d}_0) - \vartheta'_{lL+j}(\vec{d}_0) \right)^2 f(d_{l,0} | Y_1 = j) P(Y_1 = j) \int S'2(v) dv,
\end{aligned}$$

and for $l = m$ ($l, m = 1, 2, \dots, L-1$),

$$\begin{aligned}
&\Pi_{lm}(\vec{d}_0) \\
&= \lim_{n \rightarrow \infty} \frac{1}{h_n} \sum_{j=1}^L \left(\vartheta'_{(l-1)L+j}(\vec{d}_0) - \vartheta'_{lL+j}(\vec{d}_0) \right) \left(\vartheta'_{(m-1)L+j}(\vec{d}_0) - \vartheta'_{mL+j}(\vec{d}_0) \right) \\
&\quad \times E \left[S' \left(\frac{d_{l,0} - X_1}{h_n} \right) S' \left(\frac{d_{m,0} - X_1}{h_n} \right) I(Y_1 = j) \right].
\end{aligned}$$

We complete the proof of Lemma (ii).

Lemma 2.

If the conditions (C1)–(C4) hold, and nh_n^{2r+1} converges to a finite number λ as $n \rightarrow \infty$, then

$$\sqrt{nh_n} \frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \bigg|_{\vec{d} = \vec{d}_0} \xrightarrow{d} N(\lambda^{1/2} A(\vec{d}_0), \Pi(\vec{d}_0)), \quad (24)$$

where " $\xrightarrow{d}$ " denotes the convergence in distribution.

Proof of Lemma 2: Note that

$$\frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} = \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(\tilde{\mathbf{P}}_n(\vec{d})) = \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) + \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \left[\vartheta'(\tilde{\mathbf{P}}_n(\vec{d})) - \vartheta'(P(\vec{d})) \right].$$

By the similar arguments as in the proof of Theorem 1, we have

$$\sup_{\vec{d} \in \Theta} \left\| \vartheta'(\tilde{\mathbf{P}}_n(\vec{d})) - \vartheta'(\mathbf{P}_n(\vec{d})) \right\| \xrightarrow{a.s.} 0.$$

It is easy to find that

$$\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} = O_p\left(\frac{1}{nh_n}\right).$$

Thus it follows that

$$\left. \frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} = \left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) + o_p\left(\frac{1}{nh_n}\right). \quad (25)$$

Next we show the asymptotic normality of $\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0))$. According to the

results of Lemma 1, we have

$$\begin{aligned} & \sqrt{nh_n} \left[\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) - h_n^r A(\vec{d}_0) \right] \\ &= \sqrt{nh_n} \left[\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) - E \left(\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) \right) \right] \\ &+ \sqrt{nh_n} \left[E \left(\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) \right) - h_n^r A(\vec{d}_0) \right] \\ &= \sqrt{nh_n} \left[\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) - E \left(\left. \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \right|_{\vec{d} = \vec{d}_0} \vartheta'(P(\vec{d}_0)) \right) \right] + O(\sqrt{nh_n^{2r+3}}) \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[B_i(\vec{d}_0) - E(B_i(\vec{d}_0)) \right] + O(\sqrt{nh_n^{2r+3}}), \end{aligned} \quad (26)$$

where $B_i(\vec{d}_0)$ is defined in (20). Thus by the Cramér-Wold Theorem and Slutsky's Theorem, in order to obtain the asymptotic result in (24), we only need to prove that for any real vector of constants $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{L-1})^T \in \mathbb{R}^{L-1}$,

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[\lambda^T B_i(\vec{d}_0) - E(\lambda^T B_i(\vec{d}_0)) \right] \xrightarrow{d} N(0, \lambda^T \Pi(\vec{d}_0) \lambda). \quad (27)$$

Define

$$S_n = \sum_{i=1}^n \frac{w_i - E(w_i)}{(n \text{Var}(w_1))^{1/2}} \equiv \sum_{i=1}^{\tilde{n}} L_{ni}, \quad (28)$$

where $w_i \equiv w_i(\vec{d}_0) = h_n^{-1} \lambda^T B_i(\vec{d}_0)$ ($i = 1, 2, \dots, n$). Note that $\{L_{ni}\}_{i=1}^n$ is a triangular of n i.i.d random variable. We then apply the Liapunov's central limit theorem to show the asymptotic normality of S_n . Specifically, the first two conditions of Liapunov's central limit theorem are satisfied given that $E(L_{ni}) = 0$ and $\text{Var}(L_{ni}) = 1$. The third condition is also met because under condition (C1), there exists a large constant M such that

$$\begin{aligned} & \sum_{i=1}^n E|L_{ni}|^{2+\delta} \\ &= (n \text{Var}(\omega_1))^{-(1+\delta/2)} \sum_{i=1}^n E|w_i - E(w_i)|^{2+\delta} \\ &= (\text{Var}(\omega_1))^{-(1+\delta/2)} n^{-\delta/2} E|w_1 - E(w_1)|^{2+\delta} \\ &\leq (h_n \text{Var}(\omega_1))^{-(1+\delta/2)} 2^{(2+\delta)} (nh_n)^{-\delta/2} E \left| \lambda^T B_1(\vec{d}_0) \right|^{2+\delta} \\ &\leq (h_n \text{Var}(\omega_1))^{-(1+\delta/2)} 2^{(1+\delta)} (nh_n)^{-\delta/2} E \frac{1}{h_n} \|\lambda\|^2 + \delta \left\| B_1(\vec{d}_0) \right\|^{2+\delta} \\ &\leq M (h_n \text{Var}(\omega_1))^{-(1+\delta/2)} 2^{(1+\delta)} (nh_n)^{-\delta/2} \|\lambda\|^2 + \delta \sum_{k=1}^{L-1} E \frac{1}{h_n} \left| S' \left(\frac{d_{k,0} = X_1}{h_n} \right) \right|^{2+\delta}. \end{aligned}$$

Here, the first inequality follows from $E|w_1 - E(w_1)|^{2+\delta} \leq 2^{1+\delta} E|w_1|^{2+\delta}$ by the c_r inequality. From the proof of Lemma 1 (ii), $h_n \text{Var}(\omega_1) \rightarrow \lambda^T \Pi(\vec{d}_0) \lambda$. Moreover, by the nonparametric statistics theory, $E \frac{1}{h_n} \left| S' \left(\frac{d_k^0 - X_1}{h_n} \right) \right|^{2+\delta} \rightarrow \int |S'(u)|^{2+\delta} du < \infty$. All these give

$$\sum_{i=1}^n E|L_{ni}|^{2+\delta} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (29)$$

Applying the Liapunov's central limit theory to S_n gives (27). The result of Lemma 2 directly follows from combining the equations (25)–(27).

Lemma 3.

If the conditions (C1)–(C4) hold, then

$$\sup_{\vec{d} \in \Theta} \left\| \frac{\vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\frac{\partial \vec{d}}{\partial \vec{d}} \frac{\partial \vec{d}^T}{\partial \vec{d}}} - Q(\vec{d}) \right\| \xrightarrow{a.s.} 0. \quad (30)$$

Proof of Lemma 3: For $i, j = 1, 2, \dots, L$, define

$$C_{ij}(\tilde{\mathbf{P}}_n(\vec{d})) = \begin{pmatrix} \vartheta^{((i-1)L+1)((j-1)L+1)}(\tilde{\mathbf{P}}_n(\vec{d})) & \dots & \vartheta^{((i-1)L+1)(jL)}(\tilde{\mathbf{P}}_n(\vec{d})) \\ \vdots & & \vdots \\ \vartheta^{(iL)((j-1)L+1)}(\tilde{\mathbf{P}}_n(\vec{d})) & \dots & \vartheta^{(iL)(jL)}(\tilde{\mathbf{P}}_n(\vec{d})) \end{pmatrix}$$

and

$$C(\tilde{\mathbf{P}}_n(\vec{d})) = \begin{pmatrix} C_{11}(\tilde{\mathbf{P}}_n(\vec{d})) & C_{12}(\tilde{\mathbf{P}}_n(\vec{d})) & \dots & C_{1L}(\tilde{\mathbf{P}}_n(\vec{d})) \\ C_{21}(\tilde{\mathbf{P}}_n(\vec{d})) & C_{22}(\tilde{\mathbf{P}}_n(\vec{d})) & \dots & C_{2L}(\tilde{\mathbf{P}}_n(\vec{d})) \\ \vdots & \vdots & & \vdots \\ C_{L1}(\tilde{\mathbf{P}}_n(\vec{d})) & C_{L2}(\tilde{\mathbf{P}}_n(\vec{d})) & \dots & C_{LL}(\tilde{\mathbf{P}}_n(\vec{d})) \end{pmatrix}.$$

Moreover, for $i = 1, 2, \dots, L-1$, let

$$\tilde{\mathbf{H}}_i = \begin{pmatrix} \frac{1}{nh_n} \sum_{k=1}^n S' \left(\frac{d_i - X_k}{h_n} \right) I(Y_k = 1) \\ \vdots \\ \frac{1}{nh_n} \sum_{k=1}^n S' \left(\frac{d_i - X_k}{h_n} \right) I(Y_k = L) \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{H}}(\vec{d}) = \begin{pmatrix} \tilde{\mathbf{H}}_1^T & 0 & \dots & 0 \\ 0 & \tilde{\mathbf{H}}_2^T & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \tilde{\mathbf{H}}_{L-1}^T \end{pmatrix}.$$

Then after some calculations, we have

$$\begin{aligned} \frac{\vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} &= \tilde{H}(\vec{d}) \cdot I \cdot C(\tilde{\mathbf{P}}_n(\vec{d})) \cdot I^T \cdot \tilde{H}^T(\vec{d}) \\ &+ \text{diag} \left\{ \sum_{j=1}^L (\vartheta'_j(\tilde{\mathbf{P}}_n(\vec{d})) - \vartheta'_{L+j}(\tilde{\mathbf{P}}_n(\vec{d}))) \frac{1}{nh_n^2} \sum_{k=1}^n S''(\frac{d_1 - X_k}{h_n}) I(Y_k = j), \right. \\ &\quad \left. \dots, \sum_{j=1}^L (\vartheta'_j(\tilde{\mathbf{P}}_n(\vec{d})) - \vartheta'_{L+j}(\tilde{\mathbf{P}}_n(\vec{d}))) \frac{1}{nh_n^2} \sum_{k=1}^n S''(\frac{d_{L-1} - X_k}{h_n}) I(Y_k = j), \right\}, \end{aligned} \quad (31)$$

where $\text{diag}\{a_1, \dots, a_n\}$ denotes a diagonal matrix whose diagonal entries starting in the upper left corner are $a_1, \dots, a_n$. Under assumption (C1) and by similar arguments as in the proof of Theorem 1, we have

$$\sup_{\vec{d} \in \Theta} \|C(\tilde{\mathbf{P}}_n(\vec{d})) - C(P(\vec{d}))\| \xrightarrow{a.s.} 0. \quad (32)$$

By the result in [45], this implies

$$\sup_{\vec{d} \in \Theta} \|\tilde{H}(\vec{d}) - H(\vec{d})\| \xrightarrow{a.s.} 0, \quad (33)$$

and

$$\begin{aligned} \sup_{\vec{d} \in \Theta} \left| \frac{1}{nh_n^2} \sum_{k=1}^n S''(\frac{d_i - X_k}{h_n}) I(Y_k = j) - F''(d_i|Y = j)P(Y = j) \right| &\xrightarrow{a.s.} 0, \\ i = 1, 2, \dots, L-1, j = 1, 2, \dots, L. \end{aligned} \quad (34)$$

Combining (31) and (32)–(34), we conclude the proof of Lemma 3.

Proof of Theorem 2: By the Taylor series expansion, it follows that

$$0 = \frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \bigg|_{\vec{d} = \vec{d}_s} = \frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \bigg|_{\vec{d} = \vec{d}_0} + \frac{\partial^2 \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} \bigg|_{\vec{d} = \vec{d}^*} (\widehat{\vec{d}}_s - \vec{d}_0), \quad (35)$$

where $\vec{d}^*$ lies between $\widehat{\vec{d}}_s$ and $\vec{d}_0$.

We first prove (i). It follows from (25) that

$$h_n^{-r} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) \Big|_{\vec{d} = \vec{d}_0} = \frac{\partial^2 \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} \Big|_{\vec{d} = \vec{d}_*} h_n^{-r} (\widehat{\vec{d}}_s - \vec{d}_0) + o_p\left(\frac{1}{nh_n^{r+1}}\right), \quad (36)$$

By Lemma 1, we have

$$\begin{aligned} & \text{Var} \left[h_n^{-r} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) \Big|_{\vec{d} = \vec{d}_0} \right] \\ &= \frac{1}{nh_n^{2r+1}} \text{Var} \left[(nh_n)^{1/2} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) \Big|_{\vec{d} = \vec{d}_0} \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

By Chebyshev's Theorem, this leads to

$$h_n^{-r} \frac{\partial \tilde{\mathbf{P}}_n^T(\vec{d})}{\partial \vec{d}} \vartheta'(P(\vec{d})) \Big|_{\vec{d} = \vec{d}_0} \xrightarrow{P} A.$$

Coupled with Lemma 3 and equation (36), this implies the result (i).

Next we show (ii). By equation (35), it follows that

$$0 = \sqrt{nh_n} \frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \Big|_{\vec{d} = \vec{d}_0} + \frac{\partial^2 \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} \Big|_{\vec{d} = \vec{d}_*} \sqrt{nh_n} (\widehat{\vec{d}}_s - \vec{d}_0). \quad (37)$$

Thus we have

$$\begin{aligned} & \sqrt{nh_n} \left\{ \widehat{\vec{d}}_s - \vec{d}_0 + \left(\frac{\partial^2 \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} \Big|_{\vec{d} = \vec{d}_*} \right)^{-1} h_n^r A(\vec{d}_0) \right\} \\ &= \sqrt{nh_n} \left\{ \widehat{\vec{d}}_s - \vec{d}_0 + h_n^r Q^{-1}(\vec{d}_0) A(\vec{d}_0) + o_p(h_n^r) \right\} \\ &= \left(\frac{\partial^2 \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d} \partial \vec{d}^T} \Big|_{\vec{d} = \vec{d}_*} \right)^{-1} \sqrt{nh_n} \left[\frac{\partial \vartheta(\tilde{\mathbf{P}}_n(\vec{d}))}{\partial \vec{d}} \Big|_{\vec{d} = \vec{d}_0} - h_n^r A(\vec{d}_0) \right]. \end{aligned}$$

The result of (ii) follows directly from Lemma 2, Lemma 3 and Slutsky's theorem

Appendix B:: Estimation of asymptotic bias and covariance

In this section, we present the estimation of the asymptotic bias and asymptotic covariance matrix. Specifically, by the theorem proofs, the asymptotic bias and the asymptotic covariance matrix of $\widehat{\vec{d}}_s$ can be estimated respectively by

$$h_n^r \widehat{\mathcal{Q}}^{-1}(\widehat{\vec{d}}_s) \widehat{A}(\widehat{\vec{d}}_s) \quad \text{and} \quad \frac{1}{nh_n} \widehat{\mathcal{Q}}^{-1}(\widehat{\vec{d}}_s) \widehat{\Pi}(\widehat{\vec{d}}_s) \left(\widehat{\mathcal{Q}}^{-1}(\widehat{\vec{d}}_s) \right)^T,$$

Where

$$\begin{aligned} \widehat{A}(\widehat{\vec{d}}_s) &= h_n^{-r} \frac{\partial \vartheta(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s))}{\partial \widehat{\vec{d}}} \bigg|_{\widehat{\vec{d}} = \widehat{\vec{d}}_s} = \frac{1}{nh_n^{r+1}} \sum_{i=1}^n B_i^T(\widehat{\vec{d}}_s), \\ \widehat{\Pi}(\widehat{\vec{d}}_s) &= \frac{1}{nh_n} \sum_{i=1}^n B_i(\widehat{\vec{d}}_s) B_i^T(\widehat{\vec{d}}_s) \end{aligned}$$

and

$$\begin{aligned} \widehat{\mathcal{Q}}(\widehat{\vec{d}}_s) &= \widehat{H}(\widehat{\vec{d}}_s) \cdot I \cdot C(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) \cdot I^T \cdot \widehat{H}^T(\widehat{\vec{d}}_s) + \text{diag} \left(\sum_{j=1}^L \left(\vartheta'_j(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) - \vartheta'_{L+j}(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) \right) \right. \\ &\quad \times \frac{1}{nh_n^2} \sum_{k=1}^n S'' \left(\frac{\widehat{d}_{1,s} - X_k}{h_n} \right) I(Y_k = j), \dots, \sum_{j=1}^L \left(\vartheta'_j(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) - \vartheta'_{L+j}(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) \right) \\ &\quad \times \left. \frac{1}{nh_n^2} \sum_{k=1}^n S'' \left(\frac{\widehat{d}_{L-1,s} - X_k}{h_n} \right) I(Y_k = j) \right), \end{aligned}$$

With

$$B_i^T(\widehat{\vec{d}}_s) = \begin{pmatrix} \sum_{j=1}^L \left(\vartheta'_j(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) - \vartheta'_{L+j}(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) \right) S' \left(\frac{\widehat{d}_{1,s} - X_k}{h_n} \right) I(Y_k = j) \\ \vdots \\ \sum_{j=1}^L \left(\vartheta'_{(L-2)L+j}(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) - \vartheta'_{(L-1)L+j}(\widehat{\vec{P}}_n(\widehat{\vec{d}}_s)) \right) S' \left(\frac{\widehat{d}_{L-1,s} - X_k}{h_n} \right) I(Y_k = j) \end{pmatrix}.$$

$$\widehat{H}(\widehat{\vec{d}}_s) = \begin{pmatrix} \widehat{H}_1^T & 0 & \dots & 0 \\ 0 & \widehat{H}_2^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widehat{H}_{L-1}^T \end{pmatrix} \quad \text{and} \quad \widehat{H}_i = \begin{pmatrix} \sum_{k=1}^n S' \left(\frac{\widehat{d}_{i,s} - X_k}{h_n} \right) I(Y_k = 1) / nh_n \\ \vdots \\ \sum_{k=1}^n S' \left(\frac{\widehat{d}_{i,s} - X_k}{h_n} \right) I(Y_k = L) / nh_n \end{pmatrix}.$$

References

- [1]. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B, Depression, chronic diseases, and decrements in health: results from the world health surveys, *The Lancet* 370 (9590) (2007) 851–858. doi:10.1016/S0140-6736(07)61415-9.
- [2]. Zung WW, A self-rating depression scale, *Archives of general psychiatry* 12 (1) (1965) 63–70. doi: 10.1001/archpsyc.1965.01720310065008. [PubMed: 14221692]
- [3]. Altman DG, Lausen B, Sauerbrei W, Schumacher M, Dangers of using “optimal” cutpoints in the evaluation of prognostic factors, *Journal of the National Cancer Institute* 86 (11) (1994) 829–835. doi:10.1093/jnci/86.11.829. [PubMed: 8182763]
- [4]. Altman DG, Categorising continuous variables, *British journal of cancer* 64 (5) (1991) 975. doi: 10.1002/0470011815.b2a10012. [PubMed: 1931629]
- [5]. Mazumdar M, Glassman JR, Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments, *Statistics in medicine* 19 (1) (2000) 113–132. doi:10.1002/(SICI)1097-0258(20000115)19:1<113::AID-SIM245>3.0.CO;2-O. [PubMed: 10623917]
- [6]. Pepe MS, *The statistical evaluation of medical tests for classification and prediction*, Oxford University Press, USA, 2003.
- [7]. Youden WJ, Index for rating diagnostic tests, *Cancer* 3 (1) (1950) 32–35. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3. [PubMed: 15405679]
- [8]. Fluss R, Faraggi D, Reiser B, Estimation of the youden index and its associated cutoff point, *Biometrical journal* 47 (4) (2005) 458–472. doi:10.1002/bimj.200410135. [PubMed: 16161804]
- [9]. Zhou X, Qin G, Improved confidence intervals for the sensitivity at a fixed level of specificity of a continuous-scale diagnostic test, *Statistics in Medicine* 24 (3) (2005) 465–477. doi: 10.1002/sim.1563. [PubMed: 15635678]
- [10]. Schisterman EF, Faraggi D, Reiser B, Hu J, Youden index and the optimal threshold for markers with mass at zero, *Statistics in medicine* 27 (2) (2008) 297–315. doi:10.1002/sim.2993. [PubMed: 17624866]
- [11]. Zhou H, Qin G, New nonparametric confidence intervals for the youden index, *Journal of biopharmaceutical statistics* 22 (6) (2012) 1244–1257. doi:10.1080/10543406.2011.592234. [PubMed: 23075020]
- [12]. Lai C-Y, Tian L, Schisterman EF, Exact confidence interval estimation for the youden index and its corresponding optimal cut-point, *Computational statistics and data analysis* 56 (5) (2012) 1103–1114. doi:10.1016/j.csda.2010.11.023. [PubMed: 27099407]
- [13]. Wang D, Tian L, Zhao Y, Smoothed empirical likelihood for the youden index, *Computational Statistics and Data Analysis* 115 (2017) 1–10. doi:10.1016/j.csda.2017.03.014.
- [14]. Liu X, Classification accuracy and cut point selection, *Statistics in medicine* 31 (23) (2012) 2676–2686. doi:10.1002/sim.4509. [PubMed: 22307964]
- [15]. Perkins NJ, Schisterman EF, The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve, *American journal of epidemiology* 163 (7) (2006) 670–675. doi:10.1093/aje/kwj063. [PubMed: 16410346]
- [16]. Dong T, Attwood K, Hutson A, Liu S, Tian L, A new diagnostic accuracy measure and cut-point selection criterion, *Statistical methods in medical research* 26 (6) (2017) 2832–2852. [PubMed: 26486150]
- [17]. James I, Estimation of the mixing proportion in a mixture of two normal distributions from simple, rapid measurements, *Biometrics* 34 (2) (1978) 265–275. doi:10.2307/2530016. [PubMed: 667273]
- [18]. Brownie C, Habicht J-P, Selecting a screening cut-off point or diagnostic criterion for comparing prevalences of disease, *Biometrics* 40 (3) (1984) 675–684. doi:10.2307/2530910. [PubMed: 6518245]
- [19]. Baughman AL, Bisgard KM, Lynn F, Meade BD, Mixture model analysis for establishing a diagnostic cut-off point for pertussis antibody levels, *Statistics in medicine* 25 (17) (2006) 2994–3010. doi:10.1002/sim.2442. [PubMed: 16345022]

- [20]. Miller R, Siegmund D, Maximally selected chi square statistics, *Biometrics* 38 (4) (1982) 1011–1016. doi:10.1002/bimj.200510191.
- [21]. Lausen B, Schumacher M, Maximally selected rank statistics, *Biometrics* 48 (1) (1992) 73–85. doi:10.2307/2532740.
- [22]. Lausen B, Schumacher M, Evaluating the effect of optimized cutoff values in the assessment of prognostic factors, *Computational Statistics and Data Analysis* 21 (3) (1996) 307–326. doi:10.1016/0167-9473(95)00016-X.
- [23]. Betensky RA, Rabinowitz D, Maximally selected X^2 statistics for $k \times 2$ tables, *Biometrics* 55 (1) (1999) 317–320. doi:10.1111/j.0006-341X.1999.00317.x. [PubMed: 11318175]
- [24]. Hothorn T, Zeileis A, Generalized maximally selected statistics, *Biometrics* 64 (4) (2008) 1263–1269. doi:10.1111/j.1541-0420.2008.00995.x. [PubMed: 18325074]
- [25]. Peng L, Manatunga A, Wang M, Guo Y, Rahman AF, A general approach to categorizing a continuous scale according to an ordinal outcome, *Journal of statistical planning and inference* 172 (2016) 23–35. doi:10.1016/0003-4916(63)90068-X. [PubMed: 26941475]
- [26]. Brown B, Wang Y-G, Induced smoothing for rank regression with censored survival times, *Statistics in medicine* 26 (4) (2007) 828–836. doi:10.1002/sim.2576. [PubMed: 16646005]
- [27]. Heller G, Smoothed rank regression with censored data, *Journal of the American Statistical Association* 102 (478) (2007) 552–559. doi:10.1198/016214506000001257.
- [28]. Pang L, Lu W, Wang HJ, Variance estimation in censored quantile regression via induced smoothing, *Computational statistics and data analysis* 56 (4) (2012) 785–796. doi:10.1016/j.csda.2010.10.018. [PubMed: 22547899]
- [29]. Qiu Z, Qin J, Zhou Y, Composite estimating equation method for the accelerated failure time model with length-biased sampling data, *Scandinavian Journal of Statistics* 43 (2) (2016) 396–415. doi:10.1111/sjos.12182.
- [30]. Qiu Z, Wan AT, Zhou Y, Gilbert PB, Smoothed rank regression for the accelerated failure time competing risks model with missing cause of failure, *Statistica Sinica* doi:10.5705/ss.202016.0231.
- [31]. Nelder JA, Mead R, A simplex method for function minimization, *The computer journal* 7 (4) (1965) 308–313. doi:10.1093/comjnl/7.4.308.
- [32]. Nakas CT, Alonzo TA, Yiannoutsos CT, Accuracy and cut-off point selection in three-class classification problems using a generalization of the youden index, *Statistics in medicine* 29 (28) (2010) 2946–2955. [PubMed: 20809485]
- [33]. Politis RJPWM, Dimitris N, *Subsampling*, Springer, 1999.
- [34]. Kelley CT, *Iterative methods for optimization*, SIAM, 1999. doi:10.1137/1.9781611970920.
- [35]. Storn R, Price K, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization* 11 (4) (1997) 341–359. doi:10.1137/1.9781611970920.
- [36]. Price K, Storn RM, Lampinen JA, *Differential evolution: a practical approach to global optimization*, Springer Science & Business Media, 2006.
- [37]. Hansen BE, Nonparametric estimation of smooth conditional distributions, Unpublished paper: Department of Economics, University of Wisconsin doi:10.1109/TIT.2007.894631.
- [38]. Li Q, Racine JS, *Nonparametric econometrics: theory and practice*, Princeton University Press, 2007.
- [39]. Bowman A, Hall P, Prvan T, Bandwidth selection for the smoothing of distribution functions, *Biometrika* 85 (4) (1998) 799–808. doi:10.1093/biomet/85.4.799.
- [40]. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B, Depression, chronic diseases, and decrements in health: results from the world health surveys, *The Lancet* 370 (9590) (2007) 851–858. doi:10.1016/S0140-6736(07)61415-9.
- [41]. Sheehan D, Lecrubier Y, Sheehan KH, Sheehan K, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar G, The mini-international neuropsychiatric interview (m.i.n.i.): The development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10, *The Journal of Clinical Psychiatry* 59 (1998) 22–33.
- [42]. Rao CR, Mitra SK, *Generalized inverse of matrices and its applications*, Wiley New York, 1971.

- [43]. Pakes A, Pollard D, Simulation and the asymptotics of optimization estimators, *Econometrica: Journal of the Econometric Society* 57 (5) (1989) 1027–1057. doi:10.2307/1913622.
- [44]. Van Der Vaart AW, Wellner JA, *Weak Convergence and Empirical Processes*, Springer, 1996.
- [45]. Silverman BW, Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *The Annals of Statistics* 6 (1) (1978) 177–184. doi:10.1214/aos/1176345157.

Figure 1:

The plots of empirical criterion functions based on Kendall's $\tau_b(\hat{\theta}_\tau)$ (left) and MADET $(\hat{\theta}_{MD})$ (right) versus all possible cut-points for the Diabetes and Depression study data ($n = 1430$).

Table 1:

Simulation results for Scenarios 1 and 2.

Approach	Scenario	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
Non-smooth	Scenario 1	100	kendall	7.000	-0.004	-0.001	0.180	0.160	0.181	0.960	0.17
			MADET	7.000	-0.005	-0.001	0.164	0.154	0.164	0.964	0.22
			CC	7.000	-0.046	-0.007	0.163	0.155	0.169	0.933	0.06
			wkappa	7.000	-0.006	-0.001	0.164	0.154	0.164	0.964	0.08
		200	kendall	7.000	-0.000	-0.000	0.143	0.135	0.143	0.958	0.36
			MADET	7.000	-0.000	-0.000	0.128	0.127	0.128	0.967	0.37
			CC	7.000	-0.031	-0.004	0.130	0.130	0.134	0.949	0.08
			wkappa	7.000	-0.001	-0.000	0.128	0.127	0.128	0.966	1.64
		300	kendall	7.000	0.004	0.001	0.124	0.123	0.124	0.975	0.51
			MADET	7.000	0.003	0.001	0.110	0.114	0.110	0.979	0.57
			CC	7.000	-0.022	-0.003	0.112	0.116	0.114	0.965	0.115
			wkappa	7.000	0.002	0.000	0.110	0.114	0.110	0.979	2.46
	Scenario 2	100	kendall	3.641	0.001	0.000	0.631	0.618	0.631	0.951	0.12
			MADET	3.634	0.007	0.002	0.544	0.546	0.545	0.955	0.19
			CC	3.634	-0.103	-0.028	0.546	0.554	0.556	0.962	0.06
			wkappa	3.634	0.011	0.003	0.542	0.553	0.542	0.959	0.91
		200	kendall	3.641	0.004	0.001	0.521	0.509	0.521	0.932	0.46
			MADET	3.634	0.014	0.004	0.433	0.437	0.434	0.958	0.40
			CC	3.634	-0.073	-0.020	0.445	0.449	0.451	0.953	0.07
			wkappa	3.634	0.012	0.003	0.432	0.443	0.432	0.959	1.66
		300	kendall	3.641	0.016	0.004	0.452	0.452	0.452	0.953	0.53
			MADET	3.634	0.025	0.007	0.373	0.386	0.375	0.957	0.57
			CC	3.634	-0.033	-0.009	0.385	0.396	0.386	0.960	0.11
			wkappa	3.634	0.031	0.009	0.378	0.390	0.379	0.957	2.41
Smooth	Scenario 1	100	kendall	7.000	0.003	0.000	0.090	0.092	0.090	0.960	0.73
			MADET	7.000	0.003	0.000	0.092	0.100	0.092	0.966	0.80
			CC	7.000	0.004	0.001	0.100	0.100	0.100	0.955	0.73
			wkappa	7.000	0.003	0.001	0.098	0.101	0.098	0.958	0.85
		200	kendall	7.000	-0.005	-0.001	0.069	0.070	0.069	0.946	3.17
			MADET	7.000	-0.005	-0.001	0.070	0.076	0.071	0.955	3.20
			CC	7.000	-0.007	-0.001	0.075	0.074	0.076	0.947	3.08
			wkappa	7.000	-0.007	-0.001	0.074	0.075	0.075	0.948	3.24
		300	kendall	7.000	0.000	0.000	0.060	0.060	0.060	0.957	7.02
			MADET	7.000	0.001	0.000	0.062	0.065	0.062	0.963	7.56
			CC	7.000	0.001	0.000	0.065	0.066	0.065	0.955	7.09
			wkappa	7.000	0.001	0.000	0.064	0.065	0.064	0.956	7.46
	Scenario 2	100	kendall	3.641	0.144	0.040	0.331	0.302	0.361	0.899	0.69

Approach	Scenario	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
Author Manuscript		200	MADET	3.634	0.165	0.045	0.359	0.359	0.395	0.927	0.75
			CC	3.634	0.157	0.043	0.391	0.360	0.422	0.900	0.70
			wkappa	3.634	0.160	0.044	0.373	0.363	0.406	0.919	0.83
			kendall	3.641	0.080	0.022	0.261	0.246	0.273	0.922	2.60
			MADET	3.634	0.092	0.025	0.288	0.294	0.302	0.944	2.66
			CC	3.634	0.091	0.025	0.322	0.304	0.334	0.926	2.61
		300	wkappa	3.634	0.091	0.025	0.308	0.289	0.321	0.926	2.71
			kendall	3.641	0.050	0.011	0.227	0.220	0.232	0.930	5.76
			MADET	3.634	0.059	0.016	0.256	0.260	0.263	0.936	5.79
			CC	3.634	0.064	0.018	0.275	0.260	0.282	0.929	5.78
			wkappa	3.634	0.063	0.017	0.267	0.263	0.274	0.939	5.87

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 2:

Simulation results for Scenario 3.

Approach	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
Nonsmooth	100	kendall	7.000	-0.006	-0.001	0.201	0.180	0.201	0.932	2.42
			9.000	0.002	0.000	0.201	0.179	0.201	0.928	
		MADET	7.000	0.014	-0.002	0.183	0.175	0.184	0.933	4.28
			9.000	0.006	0.001	0.184	0.174	0.184	0.942	
		CC	7.000	-0.055	-0.008	0.186	0.176	0.194	0.870	0.94
			9.000	-0.046	-0.005	0.185	0.176	0.190	0.944	
		wkappa	7.000	0.030	0.004	0.187	0.176	0.189	0.951	11.15
			9.000	-0.037	-0.004	0.185	0.175	0.189	0.945	
	200	kendall	7.000	0.011	0.002	0.161	0.151	0.161	0.928	8.37
			9.000	-0.005	-0.001	0.164	0.151	0.164	0.941	
		MADET	7.000	0.003	0.000	0.148	0.144	0.148	0.917	10.19
			9.000	-0.002	-0.000	0.148	0.143	0.149	0.929	
		CC	7.000	0.002	0.000	0.090	0.087	0.090	0.936	2.79
			9.000	-0.007	-0.001	0.090	0.089	0.091	0.940	
		wkappa	7.000	0.027	0.004	0.153	0.146	0.155	0.951	39.76
			9.000	-0.025	-0.003	0.152	0.146	0.154	0.955	
	300	kendall	7.000	0.015	0.002	0.140	0.136	0.140	0.953	23.68
			9.000	-0.006	-0.001	0.142	0.137	0.143	0.946	
		MADET	7.000	0.004	0.001	0.126	0.128	0.126	0.927	22.60
			9.000	0.003	0.000	0.127	0.128	0.127	0.925	
		CC	7.000	-0.025	-0.004	0.127	0.131	0.129	0.887	7.79
			9.000	-0.033	-0.004	0.128	0.131	0.132	0.971	
		wkappa	7.000	0.028	0.004	0.130	0.130	0.133	0.967	96.47
			9.000	-0.024	-0.003	0.128	0.131	0.130	0.971	
Smooth	100	kendall	7.000	0.007	0.001	0.104	0.109	0.105	0.958	0.74
			9.000	-0.007	-0.001	0.107	0.108	0.107	0.950	
		MADET	7.000	-0.008	-0.001	0.104	0.117	0.104	0.970	0.81
			9.000	0.008	0.001	0.106	0.116	0.106	0.966	
		CC	7.000	-0.000	-0.000	0.119	0.115	0.119	0.947	0.74
			9.000	0.000	0.000	0.124	0.116	0.124	0.931	
		wkappa	7.000	0.008	0.001	0.117	0.116	0.117	0.949	0.93
			9.000	-0.008	-0.001	0.122	0.117	0.122	0.932	
	200	kendall	7.000	0.010	0.001	0.082	0.081	0.083	0.950	2.66
			9.000	-0.020	-0.002	0.082	0.081	0.085	0.938	
		MADET	7.000	-0.003	-0.000	0.083	0.087	0.083	0.959	2.80
			9.000	-0.002	-0.000	0.083	0.088	0.083	0.958	
		CC	7.000	0.006	0.001	0.089	0.087	0.089	0.942	2.72

Approach	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
		wkappa	9.000	-0.011	-0.001	0.090	0.088	0.090	0.939	
			7.000	0.006	0.001	0.089	0.087	0.089	0.942	2.74
			9.000	-0.011	-0.001	0.090	0.088	0.090	0.939	
	300	kendall	7.000	-0.000	-0.000	0.069	0.072	0.070	0.956	6.90
			9.000	-0.004	-0.000	0.071	0.070	0.071	0.944	
			7.000	-0.006	-0.000	0.071	0.076	0.071	0.961	7.12
		MADET	9.000	0.001	0.000	0.072	0.074	0.072	0.959	
			7.000	-0.001	-0.000	0.076	0.076	0.076	0.949	7.05
			9.000	-0.003	-0.000	0.078	0.075	0.078	0.944	
		wkappa	7.000	0.002	0.000	0.075	0.075	0.075	0.951	7.20
			9.000	-0.006	-0.001	0.078	0.075	0.078	0.944	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 3:

Simulation results for Scenario 4.

Approach	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
Non-smooth	100	kendall	3.762	0.005	0.001	0.765	0.705	0.765	0.938	2.71
			9.438	0.114	0.012	2.055	1.762	2.058	0.924	
		MADET	3.415	0.094	0.028	0.643	0.663	0.650	0.961	3.23
			9.948	-0.057	-0.006	1.345	1.326	1.347	0.952	
		CC	3.634	-0.089	-0.024	0.671	0.641	0.677	0.941	0.86
			9.665	-0.202	-0.021	1.441	1.429	1.456	0.948	
		wkappa	3.740	0.091	0.024	0.675	0.661	0.681	0.949	12.71
			9.295	0.066	0.007	1.355	1.362	1.357	0.954	
	200	kendall	3.762	0.004	0.001	0.598	0.571	0.598	0.941	9.55
			9.438	-0.037	-0.004	1.737	1.480	1.737	0.914	
		MADET	3.415	0.062	0.018	0.483	0.506	0.487	0.963	10.17
			9.948	-0.062	-0.006	1.115	1.067	1.117	0.945	
		CC	3.634	-0.086	-0.024	0.495	0.508	0.503	0.948	3.07
			9.665	-0.195	-0.020	1.146	1.145	1.162	0.949	
		wkappa	3.740	0.025	0.007	0.501	0.526	0.502	0.960	35.50
			9.295	0.020	0.002	1.106	1.100	1.106	0.949	
	300	kendall	3.762	0.000	0.000	0.529	0.508	0.529	0.939	23.83
			9.438	-0.005	-0.001	1.441	1.322	1.441	0.934	
		MADET	3.415	0.072	0.021	0.438	0.448	0.444	0.955	24.66
			9.948	-0.020	-0.002	0.965	0.938	0.965	0.965	
		CC	3.634	-0.042	-0.012	0.459	0.453	0.461	0.946	7.78
			9.665	-0.114	-0.012	0.991	1.000	0.998	0.949	
		wkappa	3.740	0.025	0.007	0.451	0.462	0.452	0.956	80.49
			9.295	0.035	0.004	0.949	0.955	0.950	0.947	
Smooth	100	kendall	3.762	0.219	0.058	0.397	0.403	0.453	0.914	1.27
			9.438	0.155	0.016	0.916	0.862	0.929	0.940	
		MADET	3.415	0.246	0.072	0.355	0.421	0.432	0.948	1.35
			9.948	0.306	0.031	1.114	0.943	1.155	0.897	
		CC	3.634	0.245	0.067	0.449	0.422	0.512	0.896	1.29
			9.665	0.167	0.017	1.280	0.921	1.290	0.851	
		wkappa	3.740	0.246	0.066	0.402	0.405	0.472	0.908	1.48
			9.295	0.115	0.012	1.126	0.904	1.132	0.886	
	200	kendall	3.762	0.150	0.040	0.328	0.308	0.361	0.896	2.84
			9.438	0.091	0.010	0.722	0.634	0.729	0.916	
		MADET	3.415	0.158	0.046	0.301	0.314	0.340	0.938	2.96
			9.948	0.194	0.020	0.889	0.718	0.909	0.880	
		CC	3.634	0.166	0.046	0.346	0.322	.384	0.906	2.94
			9.665	0.072	0.007	1.013	0.723	1.016	0.847	

Approach	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
	300	wkappa	3.740	0.172	0.046	0.330	0.313	0.372	0.904	3.13
			9.295	0.044	0.005	0.920	0.690	0.921	0.845	
		kendall	3.762	0.099	0.026	0.278	0.262	0.295	0.918	5.65
			9.438	0.074	0.008	0.645	0.562	0.649	0.909	
		MADET	3.415	0.117	0.034	0.256	0.270	0.282	0.936	5.84
			9.948	0.134	0.013	0.771	0.620	0.783	0.882	
		CC	3.634	0.119	0.033	0.276	0.274	0.301	0.921	5.61
			9.665	0.075	0.008	0.829	0.615	0.832	0.839	
		wkappa	3.740	0.118	0.032	0.264	0.271	0.289	0.927	5.93
			9.295	0.051	0.005	0.779	0.591	0.781	0.855	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs

Table 4:

Simulation results for Scenario 5.

n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
100	kendall	7.000	0.016	0.000	0.119	0.122	0.120	95.2%	0.82
		9.000	0.001	0.000	0.117	0.127	0.117	96.2%	
		11.000	-0.013	-0.000	0.112	0.119	0.113	95.8%	
	MADET	7.000	-0.023	-0.000	0.119	0.130	0.121	96.6%	0.99
		9.000	-0.000	-0.000	0.113	0.131	0.113	97.1%	
		11.000	0.026	0.000	0.110	0.126	0.113	96.8%	
	CC	7.000	-0.004	-0.001	0.136	0.128	0.136	93.4%	0.87
		9.000	0.004	0.000	0.136	0.130	0.136	94.0%	
		11.000	0.003	0.000	0.129	0.126	0.129	94.9%	
	wkappa	7.000	0.018	0.003	0.134	0.130	0.135	94.0%	1.05
		9.000	0.004	0.000	0.135	0.131	0.135	94.3%	
		11.000	-0.019	-0.002	0.127	0.126	0.129	95.1%	
200	kendall	7.000	0.013	0.000	0.089	0.092	0.090	95.3%	2.76
		9.000	-0.001	-0.000	0.092	0.093	0.092	95.2%	
		11.000	-0.018	-0.000	0.092	0.092	0.093	93.9%	
	MADET	7.000	-0.017	-0.002	0.088	0.096	0.090	95.9%	2.98
		9.000	-0.001	-0.000	0.090	0.097	0.090	96.3%	
		11.000	0.011	0.001	0.091	0.097	0.092	95.8%	
	CC	7.000	0.001	0.000	0.101	0.096	0.101	94.0%	2.84
		9.000	-0.001	-0.000	0.098	0.097	0.098	95.5%	
		11.000	-0.004	-0.000	0.100	0.097	0.100	94.0%	
	wkappa	7.000	0.018	0.003	0.101	0.096	0.103	93.3%	3.17
		9.000	-0.001	-0.000	0.097	0.097	0.097	95.6%	
		11.000	-0.021	-0.002	0.100	0.097	0.102	93.5%	
300	kendall	7.000	0.011	0.000	0.077	0.077	0.078	94.3%	6.20
		9.000	-0.001	-0.000	0.078	0.081	0.078	95.8%	
		11.000	-0.014	-0.000	0.076	0.079	0.078	95.6%	
	MADET	7.000	-0.017	-0.002	0.077	0.082	0.079	95.4%	6.29
		9.000	-0.001	-0.000	0.077	0.084	0.077	96.9%	
		11.000	0.011	0.001	0.076	0.084	0.077	96.4%	
	CC	7.000	-0.001	-0.000	0.085	0.081	0.085	94.4%	6.40
		9.000	0.000	0.000	0.085	0.085	0.085	94.7%	
		11.000	-0.001	-0.000	0.083	0.083	0.083	95.0%	
	wkappa	7.000	0.018	0.003	0.085	0.082	0.087	93.7%	6.74
		9.000	-0.001	-0.000	0.084	0.085	0.084	94.9%	
		11.000	-0.021	-0.002	0.083	0.083	0.085	94.3%	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 5:

Simulation results for Scenario 6.

n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
100	kendall	3.812	0.291	0.076	0.530	0.462	0.605	89.1%	1.55
		9.677	0.255	0.026	1.309	0.964	1.333	85.5%	
		19.504	0.415	0.021	2.239	1.128	2.277	67.7%	
	MADET	3.383	0.294	0.087	0.422	0.484	0.515	92.6%	1.61
		9.628	0.407	0.042	1.291	1.008	1.353	87.5%	
		21.087	0.266	0.013	2.431	1.045	2.446	60.8%	
	mc	3.634	0.294	0.081	0.518	0.485	0.596	89.7%	1.50
		9.665	0.195	0.020	1.446	1.017	1.459	84.9%	
		20.059	0.311	0.015	2.496	1.109	2.515	62.8%	
	wkappa1	3.762	0.285	0.076	0.472	0.464	0.551	90.8%	1.77
		9.618	0.164	0.017	1.306	1.136	1.316	91.0%	
		19.544	0.340	0.016	2.307	1.183	2.328	69.3%	
200	kendall	3.812	0.226	0.059	0.341	0.350	0.409	90.0%	4.87
		9.677	0.164	0.017	0.912	0.890	0.926	91.1%	
		19.504	0.208	0.011	1.665	1.069	1.678	80.0%	
	MADET	3.383	0.238	0.070	0.288	0.339	0.373	92.5%	5.05
		9.628	0.333	0.035	0.945	0.853	1.002	91.0%	
		21.087	0.193	0.009	1.822	1.033	1.832	71.3%	
	mc	3.634	0.244	0.067	0.353	0.350	0.429	89.0%	4.73
		9.665	0.184	0.019	1.101	0.871	1.116	88.3%	
		20.059	0.161	0.008	1.812	1.048	1.819	75.3%	
	wkappa1	3.762	0.246	0.065	0.333	0.338	0.413	89.3%	5.00
		9.618	0.142	0.015	0.957	0.814	0.968	89.8%	
		19.544	0.112	0.006	1.671	1.010	1.675	77.7%	
300	kendall	3.812	0.192	0.050	0.294	0.301	0.351	90.4%	7.91
		9.677	0.177	0.018	0.803	0.710	0.823	90.6%	
		19.504	0.269	0.014	1.473	0.980	1.498	79.9%	
	MADET	3.383	0.200	0.059	0.247	0.279	0.318	91.6 %	7.72
		9.628	0.288	0.030	0.822	0.749	0.871	91.1%	
		21.087	0.279	0.013	1.585	0.961	1.610	75.5%	
	mc	3.634	0.201	0.055	0.298	0.299	0.360	90.8%	7.49
		9.665	0.145	0.015	0.909	0.734	0.920	88.7%	
		20.059	0.235	0.012	1.625	0.937	1.642	73.5%	
	wkappa1	3.762	0.202	0.054	0.281	0.282	0.346	89.4%	8.06
		9.618	0.141	0.015	0.823	0.701	0.835	89.8%	
		19.544	0.211	0.011	1.509	0.959	1.524	77.9%	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 6:

Simulation results for Scenarios 7 and 8.

Scenario	n	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
Scenario 7	300	kendall	7.037	0.012	0.002	0.113	0.105	0.113	92.7%	6.860
			8.975	-0.018	-0.002	0.110	0.101	0.112	91.8%	
		MADET	6.991	0.000	0.000	0.112	0.109	0.112	93.4%	7.171
			9.009	-0.008	-0.001	0.110	0.108	0.111	93.7%	
		CC	7.250	0.036	0.005	0.112	0.106	0.117	92.6%	6.940
			8.710	-0.037	-0.004	0.104	0.103	0.110	93.0%	
		wkappa	7.252	0.037	0.005	0.111	0.106	0.117	92.5%	7.018
			8.710	-0.037	-0.004	0.103	0.102	0.110	92.6%	
Scenario 8	300	kendall	4.041	0.078	0.019	0.417	0.363	0.424	90.5%	5.427
			9.186	0.020	0.002	0.877	0.637	0.877	83.9%	
		MADET	3.415	0.138	0.040	0.411	0.394	0.433	92.5%	5.645
			9.948	0.092	0.009	1.079	0.619	1.083	75.3%	
		CC	5.022	0.130	0.026	0.481	0.977	0.498	100 %	5.463
			6.145	0.019	0.003	0.732	0.829	0.732	96.8%	
		wkappa	4.686	0.116	0.025	0.383	0.356	0.400	92.1%	5.598
			6.613	0.019	0.003	0.616	0.492	0.617	88.4%	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 7:

Simulation results for Scenario 3 with $n = 100$ and 200 , and $S(\cdot)$ set as the integral of fourth order of Epanechnikov kernel ($r = 4$), second order of Gaussian kernel ($r = 2$), or fourth order of Gaussian kernel ($r = 4$).

n	Kernel	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
100	Epanechnikov, $r = 4$	kendall	7.000	0.001	0.000	0.113	0.109	0.113	94.8%	1.24
			9.000	-0.001	-0.000	0.118	0.108	0.118	91.8%	
		MADET	7.000	-0.004	-0.001	0.098	0.109	0.098	96.8%	1.31
			9.000	0.005	0.001	0.101	0.108	0.101	95.4%	
		CC	7.000	0.001	0.000	0.113	0.109	0.113	94.8%	1.24
			9.000	-0.001	-0.000	0.118	0.108	0.118	91.8%	
		wkappa	7.000	0.008	0.001	0.112	0.110	0.112	95.2%	1.40
			9.000	-0.008	-0.001	0.117	0.108	0.117	92.2%	
	Gaussian, $r = 2$	kendall	7.000	0.006	0.001	0.105	0.109	0.105	95.7%	0.66
			9.000	-0.007	-0.001	0.107	0.107	0.108	94.7%	
		MADET	7.000	-0.009	-0.001	0.104	0.116	0.104	97.0%	0.73
			9.000	0.008	0.001	0.107	0.115	0.107	95.7%	
		CC	7.000	-0.001	-0.000	0.120	0.116	0.120	95.3%	0.67
			9.000	-0.000	-0.000	0.124	0.115	0.124	92.4%	
		wkappa	7.000	0.007	0.001	0.117	0.116	0.118	96.0%	0.81
			9.000	-0.008	-0.001	0.122	0.114	0.123	92.7%	
	Gaussian, $r = 4$	kendall	7.000	0.008	0.001	0.098	0.102	0.098	95.7%	0.86
			9.000	-0.008	-0.001	0.102	0.102	0.102	94.2%	
		MADET	7.000	-0.005	-0.001	0.094	0.108	0.095	97.4%	0.91
			9.000	0.005	0.001	0.099	0.108	0.099	95.8%	
		CC	7.000	0.000	0.000	0.111	0.108	0.111	95.4%	0.85
			9.000	-0.000	-0.000	0.116	0.108	0.116	92.5%	
		wkappa	7.000	0.009	0.001	0.109	0.108	0.110	95.4%	1.02
			9.000	-0.009	-0.001	0.116	0.108	0.116	92.3%	
200	Epanechnikov, $r = 4$	kendall	7.000	0.002	0.000	0.076	0.075	0.076	94.4%	4.73
			9.000	-0.009	-0.001	0.075	0.076	0.075	94.9%	
		MADET	7.000	-0.001	-0.000	0.075	0.079	0.075	95.6%	4.80
			9.000	-0.004	-0.000	0.074	0.080	0.074	96.3%	
		CC	7.000	0.002	0.000	0.084	0.079	0.084	93.8%	4.78
			9.000	-0.007	-0.000	0.082	0.081	0.082	93.9%	
		wkappa	7.000	0.004	0.001	0.083	0.079	0.083	93.9%	4.92
			9.000	-0.009	-0.001	0.081	0.080	0.082	93.8%	
	Gaussian, $r = 2$	kendall	7.000	0.005	0.001	0.083	0.082	0.084	95.2%	2.43
			9.000	-0.010	-0.001	0.083	0.083	0.084	94.6%	
		MADET	7.000	-0.003	-0.000	0.084	0.087	0.084	96.0%	2.54

n	Kernel	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
		CC	9.000	-0.002	-0.000	0.084	0.088	0.084	95.9%	
			7.000	0.003	0.000	0.091	0.087	0.091	94.1%	2.42
		wkappa	9.000	-0.007	-0.001	0.091	0.088	0.091	93.5%	
			7.000	0.006	0.001	0.090	0.087	0.090	94.4%	2.58
			9.000	-0.011	-0.001	0.090	0.088	0.091	93.5%	
	Gaussian, $r = 4$	kendall	7.000	0.003	0.000	0.074	0.074	0.074	95.0%	2.58
			9.000	-0.009	-0.001	0.073	0.075	0.074	94.7%	
		MADET	7.000	-0.002	-0.000	0.074	0.078	0.074	96.0%	3.65
			9.000	-0.004	-0.000	0.073	0.079	0.073	96.0%	
		CC	7.000	0.002	0.000	0.082	0.078	0.082	94.3%	3.52
			9.000	-0.007	-0.001	0.081	0.079	0.081	93.5%	
		wkappa	7.000	0.005	0.001	0.081	0.078	0.081	94.6%	3.76
			9.000	-0.010	-0.001	0.080	0.079	0.081	93.8%	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 8:

Simulation results for Scenario 4 with $n = 100$ and 200 , and $S(\cdot)$ set as the integral of fourth order of Epanechnikov kernel ($r = 4$), second order of Gaussian kernel ($r = 2$), or fourth order of Gaussian kernel ($r = 4$).

n	Kernel	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
100	Epanechnikov, $r=4$	kendall	3.762	0.045	0.012	0.514	0.407	0.516	88.5%	2.30
			9.438	0.107	0.011	1.122	0.624	1.128	72.2%	
		MADET	3.415	0.109	0.032	0.505	0.456	0.516	92.1%	2.28
			9.948	0.113	0.011	1.321	0.618	1.326	64.7%	
		CC	3.634	0.060	0.011	0.558	0.435	0.562	88.3%	2.15
			9.665	0.108	0.011	1.449	0.630	1.453	62.1%	
		wkappa	3.740	0.043	0.012	0.525	0.452	0.527	91.4%	2.46
			9.295	0.087	0.009	1.305	0.614	1.308	64.5%	
	Gaussian, $r=2$	kendall	3.762	0.223	0.059	0.399	0.414	0.457	91.9%	1.03
			9.438	0.161	0.017	0.933	0.867	0.947	93.2%	
		MADET	3.415	0.244	0.071	0.371	0.430	0.444	93.8%	1.07
			9.948	0.304	0.031	1.122	0.972	1.163	90.1%	
		CC	3.634	0.245	0.067	0.459	0.431	0.520	89.9%	1.03
			9.665	0.180	0.019	1.299	0.948	1.312	85.0%	
		wkappa	3.740	0.248	0.066	0.407	0.416	0.476	91.1%	1.18
			9.295	0.113	0.012	1.136	0.921	1.142	88.5%	
	Gaussian, $r=4$	kendall	3.762	0.068	0.018	0.466	0.434	0.471	93.0%	1.46
			9.438	0.118	0.013	1.091	0.725	1.097	80.8%	
		MADET	3.415	0.129	0.038	0.445	0.441	0.463	93.6%	1.53
			9.948	0.123	0.012	1.256	0.761	1.262	76.9%	
		CC	3.634	0.085	0.024	0.560	0.446	0.517	91.5%	1.42
			9.665	0.117	0.012	1.408	0.768	1.413	72.6%	
		wkappa	3.740	0.067	0.018	0.474	0.436	0.479	92.8%	1.53
			9.295	0.094	0.010	1.271	0.737	1.274	75.5%	
200	Epanechnikov, $r=4$	kendall	3.762	0.027	0.007	0.378	0.332	0.379	90.9%	5.44
			9.438	0.051	0.005	0.840	0.592	0.841	82.9%	
		MADET	3.415	0.054	0.016	0.350	0.346	0.354	94.7%	5.50
			9.948	0.040	0.004	0.984	0.643	0.985	80.4%	
		CC	3.634	0.035	0.010	0.395	0.358	0.397	92.9%	5.43
			9.665	0.017	0.002	1.117	0.631	1.117	74.1%	
		wkappa	3.740	0.032	0.009	0.387	0.358	0.388	92.5%	5.55
			9.295	0.023	0.003	1.033	0.647	1.033	76.6%	
	Gaussian, $r=2$	kendall	3.762	0.155	0.041	0.323	0.311	0.358	90.1%	1.55
			9.438	0.094	0.010	0.722	0.638	0.728	91.7%	
		MADET	3.415	0.167	0.049	0.301	0.313	0.344	92.6%	1.62

n	Kernel	Criterion	$\vec{d}_0$	BIAS	RBIAS	SE	ASE	RMSE	CP	CT
		CC	9.948	0.201	0.020	0.886	0.773	0.909	89.9%	
			3.634	0.174	0.048	0.344	0.323	0.385	89.8%	1.49
		wkappa	9.665	0.073	0.008	1.020	0.778	1.023	87.8%	
			3.740	0.178	0.048	0.325	0.312	0.370	90.4%	1.63
			9.295	0.057	0.006	0.926	0.713	0.928	86.3%	
	Gaussian,r=4	kendall	3.762	0.052	0.014	0.340	0.328	0.344	93.2%	4.57
			9.438	0.058	0.006	0.802	0.635	0.804	87.7%	
		MADET	3.415	0.085	0.025	0.315	0.324	0.327	95.0%	4.54
			9.948	0.064	0.006	0.959	0.725	0.962	87.1%	
		CC	3.634	0.065	0.018	0.357	0.335	0.363	93.3%	4.56
			9.665	0.019	0.002	1.079	0.731	1.080	83.3%	
		wkappa	3.740	0.059	0.016	0.345	0.333	0.350	93.8%	4.58
			9.295	0.028	0.003	1.003	0.679	1.003	79.3%	

BIAS: empirical bias; RBIAS: empirical relative bias; SE: empirical standard deviations; ASE: average estimated standard deviation; RMSE: root mean square error; CP: empirical coverage probability of 95% confidence interval; CT: total computational time (in hour) for 1000 simulation runs.

Table 9:

Results from analyzing the Diabetes and Depression Study with three depression categories. EST: estimated cut-point; SE: standard error of cut-point estimate; CI: 95% confidence interval; CTs: total computational time (in seconds).

Criterion	Cut-point	<i>Non – smooth</i>				<i>Smooth</i>			
		EST	SE	CI	CTs	EST	SE	CI	CTs
kendall	$\hat{d}_1$	43.40	1.26	(40.89, 45.91)	24.50	42.12	1.14	(39.88,44.35)	26.50
	$\hat{d}_2$	57.04	1.17	(54.54, 59.53)		55.88	0.79	(54.32,57.44)	
MADET	$\hat{d}_1$	41.88	3.06	(35.23, 48.51)	63.65	41.54	0.83	(39.92,43.16)	28.78
	$\hat{d}_2$	61.88	3.01	(54.63, 69.12)		61.72	2.11	(57.59,65.85)	
mc	$\hat{d}_1$	56.41	1.14	(54.22, 58.60)	8.17	57.31	0.78	(55.77,58.84)	24.75
	$\hat{d}_2$	57.04	1.26	(54.70, 59.37)		57.31	1.42	(54.52,60.09)	
wkappa1	$\hat{d}_1$	55.76	1.41	(52.67, 58.84)	183.05	54.93	0.76	(53.45,56.41)	26.66
	$\hat{d}_2$	57.04	1.16	(54.70, 59.37)		56.58	0.89	(54.84,58.32)	

Table 10:

Results from analyzing the Diabetes and Depression Study with two depression categories. Results from analyzing the Diabetes and Depression Study with two depression categories. EST: estimated cut-points; SE: standard error of $\hat{d}_s$; CI: 95% confidence interval; CTs: total computational time (in seconds).

Criterion	Cut-point	<i>Non – smooth</i>				<i>Smooth</i>			
		EST	SE	CI	CTs	EST	SE	CI	CTs
kendall	$\hat{d}_1$	49.38	1.66	(46.29, 52.46)	1.38	51.50	0.66	(50.21,52.79)	18.78
MADET	$\hat{d}_1$	49.38	1.86	(45.68, 53.07)	2.59	51.51	0.79	(49.96,53.06)	19.50
mc	$\hat{d}_1$	57.69	1.49	(54.88, 60.51)	1.01	58.76	0.68	(57.34,60.00)	24.50
wkappa1	$\hat{d}_1$	57.00	1.35	(53.34, 60.73)	18.39	56.31	0.76	(54.82,57.80)	17.42