sciendo

Which of the favorite optimal cut-off determination methods is preferable for the ordinal response data? A simulation study

Afra ALKAN
Ankara Yildirim Beyazit University, Ankara, Turkey
afra.alkan@gmail.com

Selcen YÜKSEL
Ankara Yildirim Beyazit University, Ankara, Turkey
selcenpehlivan@gmail.com

Pervin DEMİR
Ankara Yildirim Beyazit University, Ankara, Turkey
pervin.demr@gmail.com

Abstract. *The aim of this study is to investigate the performance of the optimal cut-off methods, which are generally used for the diagnostic tests with the continuous response, for the tests with the ordinal response. Diagnostic accuracy studies examine the ability of a diagnostic test to discriminate between the patients with and without the condition. For diagnostic tests with a continuous response, it is important in practice to calculate the optimal cut-off point that can differentiate patients and healthy individuals. There are many methods proposed in the literature to obtain the optimal cut-point value for continuous test results. The Youden index, the point closest-to-(0, 1) corner in the ROC plane approach, the concordance probability, and the minimum P-value approach are commonly used methods to determine optimal-cut-point. But the researches examining the performance of these methods in the setting of the ordinal response tests are lacking in the literature. So, we compared the mentioned optimal cut-off methods for the ordinal response data by the way of simulation design by considering the sample size and the balance of groups as simulation conditions. The sample sizes of the diseased and non-diseased group were set (50, 50), (100, 100), and (200, 200) for balanced design and (50, 100), (50, 150) and (50, 200) for unbalanced design. For each scenario, 1000 repeats were generated. The differences between the estimated and the true cut-off points (biases) were calculated. All these methods overestimated the true cut-off point, but the median biases of the methods were varying. For the unbalanced design, the same result was relevant but for the balanced design, the minimum P-value approach had a median bias as 0 while others have 1.*

Keywords: ordinal data, optimal cut-off, Youden index, minimum P-value, concordance probability, point closest-to-(0, 1) corner in the ROC plane.

Introduction

Diagnostic medicine is the process of identifying the disease, or condition, that a patient has, and ruling out conditions that the patient does not have, through assessment of the patient's signs, symptoms, and results of various diagnostic tests. Diagnostic accuracy studies are research studies that examine the ability of diagnostic tests to discriminate between patients with and without the condition (Zhou et al., 2011).

In the case of a numeric or ordinal response test, it is the interest to determine the optimal cut-off point which is used to classify subjects as testing positive from those testing negative. The receiver operating characteristic curve (ROC) is often the starting point for determining the optimal cut-point (Rota and Antolini, 2014). The ROC curve is a plot of the true positive fraction (TPF) and the false positive fraction (FPF) for all possible cut-point values of the test (Pepe, 2003). The selection of the cut-off point necessitates a compromise between sensitivity and specificity (Liu, 2012; Habibzadeh et al, 2016). Although there are some criteria that consider both sensitivity and specificity, the optimal cut-off point is criterion dependent (Liu, 2012).

There are several methods to determine the optimal cut-off point. The Youden index (Youden, 1950), the point closest-to-(0, 1) corner in the ROC plane approach (Perkins and Schisterman, 2006), the concordance probability (Liu, 2012), the minimum P-value approach (Miller and Siegmund, 1982) are the most known and commonly used methods for diagnostics tests with continuous response.

According to the distribution of the continuous test result or the proportion of the number of individuals in the patient and healthy group, there are some simulation studies to search which of these tests are superior in determining the optimal cut-off point (Rota and Antolini, 2014). But the researches examining the performance of these methods in the setting of the ordinal response tests are lacking in the literature. Thus, the aim of this study is to investigate the performance of the optimal cut-off methods, which are generally used for diagnostic tests with a continuous response, for tests with an ordinal response.

Literature review

The contents of the mentioned methods are given below.

Let X be an ordinal test with five possible results which is assumed to be related to the true disease status, where D and $\bar{D}$ present the presence and the absence of the disease, respectively. The true positive fraction TPF(c) and the false positive fraction FPF(c) are respectively defined, at any given possible cut-off point c of X, as

$$TPF(c) = P(X > c|D) = S_D(c)$$

and

$$FPF(c) = P(X > c|\bar{D}) = S_{\bar{D}}(c)$$

The minimum P-value approach (minP):

The minimum P-value approach (Miller and Siegmund, 1982) is based on a systematic search of the optimal cut-off point that achieves the minimum of the P-value of the Chi-square test statistic on the absence of association between the resulting dichotomized biomarker and the binary outcome, or, in other words, the maximum of the associated Chi-square statistic over all possible cut-off point values c of X. The Chi-square objective function is

$$CHI_1^2(c) = \frac{\left(S_D(c) - S_{\bar{D}}(c)\right)^2}{\left(\frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}}\right)\left(1 - \frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}}\right)\left(\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}\right)}$$

where n_D is the number of diseased subjects and $n_{\bar{D}}$ is the number of non-diseased subjects (Rota and Antolini, 2014).

Youden Index method (J):

The Youden index (J) (Youden, 1950) is the maximum achievable value of the Youden function J(c), defined as the difference between the population quantities TPF(c) and FPF(c)

$$J(c) = S_D(c) - S_{\bar{D}}(c)$$

The optimal cut-off point $\hat{c}_J$ is the c that achieves the maximum of the Youden Function $\hat{J}(c)$ over all possible cut-off values of X.

Concordance probability method (CZ):

The concordance probability (Liu, 2012) objective function could be defined as the product of the population quantities TPF(c) and the complement to one of FPF(c),

$$CZ(c) = S_D(c)\big(1 - S_{\bar{D}}(c)\big)$$

The optimal cut-off point $\hat{c}_{CZ}$ is the c that achieves the maximum of the concordance probability function $\hat{C}Z(c)$ over all possible cut-off values of X.

Point closest-to-(0, 1) corner in the ROC plane approach (ER):

The objective function of this ROC-based method (Perkins and Schisterman, 2006) could be easily defined by applying the Euclidean distance formula between the point on the ROC plane defined by the population quantities TPF(c) and FPF(c) and the point (0, 1),

$$ER(c) = \sqrt{S_{\bar{D}}(c)^2 + (S_D(c) - 1)^2}$$

The optimal cut-off point $\hat{c}_{ER}$ is the c that achieves the minimum of the objective function $\widehat{ER}(c)$ over all possible cut-off values of X.

We assumed that X has five possible ordinal outcomes as 0: definitely negative, 1: probably negative, 2: suspicious, 3: probably positive and 4: definitely positive for the presence of the disease. Data were generated using item response theory to ensure that higher responders were more likely to be diseased. So, the rating scale model (RSM) was used in this study because responses were given on a 0-4-point rating scale, thought that with the same threshold.

Let θ be the underlying latent trait related to what test measures and β be the item difficulty. The RSM is

$$\boldsymbol{\pi_{nc}} = \frac{\boldsymbol{exp\sum_{j=0}^{c}\Big(\theta_n - \big(\beta + \tau_j\big)\Big)}}{\boldsymbol{\sum_{k=0}^{4} exp\sum_{j=0}^{k}\Big(\theta_n - \big(\beta + \tau_j\big)\Big)}} \quad \boldsymbol{j = 0, 1, 2, 3} \tag{1}$$

$$\boldsymbol{\tau_j \equiv 0, exp\sum_{j=0}^{0}(\theta_n - (\beta + \tau_j)) = 1}$$

where π_{nc} is the probability of resulting in a score of c for individual n, β is the item difficulty and τ_j is the threshold of the jth category (Andrich, 1978).

Let us assume that the latent trait is normally distributed for non-diseased and diseased populations, respectively as $\theta_{\bar{D}} \sim N(\mu_{\bar{D}} = 0, \sigma_{\bar{D}}^2 = 1)$ and $\theta_D \sim N(\mu_D, \sigma_D^2 = 1)$ (Figure 1). These two distributions intersect at $\mu_D/2$ resulting in the optimal cut-off point for the latent trait to discriminate the diseased subjects from those without the disease (Rota and Antolini, 2014).

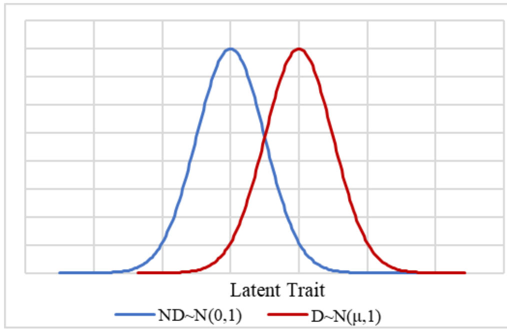

Figure 1. The distribution of the latent trait in disease and non-diseases population (ND: Non-diseased population, D: Diseased population)

The test result of a person with the latent trait of $\mu_D/2$ is estimated by equation 1 with specified item parameters, which gives the optimal cut-off point for the ordinal response test.

Methodology

Simulation design

R language (ver. 3.5) and RStudio (Version 1.1.463 – © 2009-2018 RStudio, Inc) were used for simulation (R Core Team, 2018). We considered the balanced and unbalanced designs for 1000 repeats for each scenario. The sample sizes of the diseased and non-diseased samples were set 50, 100 and 200 in the balanced design, and (50, 100), (50, 150) and (50, 200) in the unbalanced design.

We first randomly generated the latent trait of non-diseased samples from N(0,1). For the diseased samples, μ_D is set to equal {0.51, 1.05, 1.68, 2.56} resulting the optimum cut-off points of latent trait as {0.255, 0.575, 0.84, 1.28}. Then, we set the item difficulty of the ordinal test as 0.25 and the category thresholds as {-2.25, -0.75, 0.75, 2.25}. The test score probabilities of a person with the latent trait of {0.255, 0.575, 0.84, 1.28} were estimated by substituting the difficulty and the category thresholds in equation 1. The test score with the highest response probability was considered as the true cut-off points of the test, which are {2, 2, 2, 3}.

We generated the test results of the samples from RSM via the *genPattern()* function of the *catR* package (Magis D. and Raiche G., 2012) by using the latent trait and the item parameters.

After data generation and determination of the optimal true cut-off points, the cut-off points maximizing the functions of J, CHI^2, CZ, and minimizing the ER function were estimated via the *optimal.cutpoints()* function of the *OptimalCutpoints* package (Lopez-Raton M. et al.,2014). In the case of multiple values satisfying the corresponding conditions, the minimum of these values was selected as the estimated cut-off point. The bias of each method was defined as the difference between the estimated cut-off points and the corresponding true cut-off point. The same procedure was followed in the case of the unbalanced design. The design of the simulation study is summarized in the flow chart given below (Figure 2).

Sample size: The sample sizes of diseased and non-diseased samples are (50, 50); (100, 100); (200, 200) in balanced designs; (50, 100); (50, 150); (50, 200) in unbalanced designs were set as simulation condition.

The latent trait was generated by $N(\mu = 0, \sigma^2 = 1)$ for non-diseased samples and by $N(\mu_D, \sigma^2 = 1)$ for diseased samples; μ_D={0.51, 1.05, 1.68, 2.56}. $\mu_D/2$ is the intersection point of the latent trait (Rota and Antolini, 2014)?

The difficulty and the ordered response thresholds of the ordinal test were set to equal 0.25 (moderate difficulty), and {-2.25, -0.75, 0.75, 2.25}, respectively.

The test scores of the samples were obtained from RSM with these item parameters.

The intersection points of the latent trait are {0.255, 0.575, 0.84, 1.28} logits. The probabilities of each test score for these logits were estimated with RSM. The score that has the maximum probability was determined as the true cut-off value.

The true cut-off points for ordinal response test are obtained as {2, 2, 2, 3}.

The cut-off point values for the generated data were estimated by the minP, J, CZ and ER methods via the related R packages.

The difference between the true cut-off value and the estimated cut-off point is calculated for each method. All steps were repeated 1000 times.

Figure 2. The steps of the simulation study

Results and discussions

The range of bias was the widest in the J method and the narrowest in the ER method, while all methods had the same median bias of 1 (Figure 3). This result didn't change in the unbalanced design (Figure 4). The median bias of the minP method decreased to 0 and the range of bias in CZ method became narrower in the balanced design, however, the bias of other methods remained the same (Figure 4).

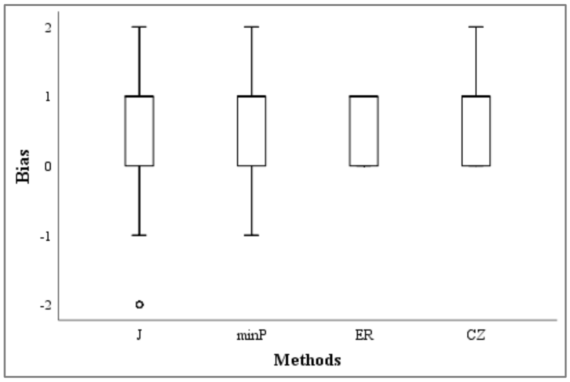

Figure 3. The bias distribution of each method

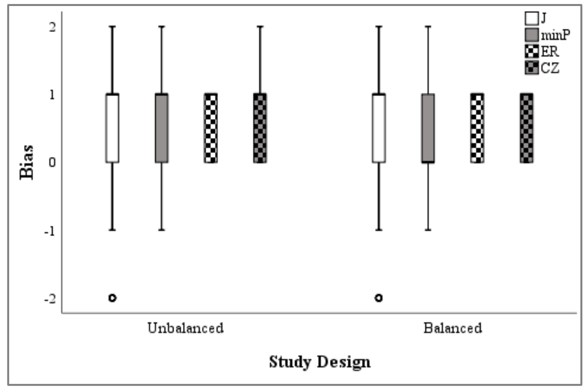

Figure 4. The bias distribution of methods based on the study design

Table 1 shows the descriptive statistics of the bias based on the sample size in the balanced design. The median bias of the J method decreased to 0 when the sample size was 50. For other methods, the sample size didn't affect the distribution of the bias.

Table 1. Median (min; max) bias for the methods-balanced scenario.

Sample size $n_D = n_{\bar{D}}$	J	minP	ER	CZ
50	0(-2;2)	0(-1;2)	1(0;1)	1(0;1)
100	1(-2;2)	0(-1;2)	1(0;1)	1(0;1)
200	1(-1;2)	0(-1;2)	1(0;1)	1(0;1)
Overall	1(-2;2)	0(-1;2)	1(0;1)	1(0;1)

In the unbalanced design, all methods had the median bias of 1 in all sample sizes. The sample size didn't change the bias distribution in all methods except CZ (Table 2). When the sample size in the non-diseased group was 100 and 150, the range of the bias got narrower for the CZ method. The ER method outperformed compared to the other methods considering the range of bias.

Table 2. Median (min; max) bias for the methods-unbalanced scenario.

Sample size		J	minP	ER	CZ
n_D	$n_{\bar{D}}$				
50	100	1(-2;2)	1(-1;2)	1(0;1)	1(0;1)
50	150	1(-2;2)	1(-1;2)	1(0;1)	1(0;1)
50	200	1(-2;2)	1(-1;2)	1(0;1)	1(0;2)
Overall		1(-2;2)	1(-1;2)	1(0;1)	1(0;2)

Since the studies examining the performance of these methods for the ordinal response tests are lacking in the literature, we cannot compare our results. But, we can say that our results are similar to those of Rota and Antolini (2014) irrespective of the data type of the diagnostic test.

Conclusion

In this study, the simulation design was conducted under similar conditions with Rota and

Antolini (2014). There were also the distributions of the continuous test among the simulation conditions of that study. For all simulation scenarios they held, the point closest to (0-1) corner in the ROC plane and the concordance probability approaches showed a better performance in the estimation of the optimal cut-off, compared to the minimum P-value and Youden Index methods. In our study, similar methods were investigated for diagnostic tests with ordinal responses, unlike the Rota's study. When the range of bias in both balanced and unbalanced scenarios are considered, the point closest to (0-1) corner in the ROC plane and the concordance probability methods are optimistic, but when the median biases are examined, it can be said that the minimum P-value method is better than others in the balanced scenario.

References

Andrich D. (1978). "A rating formulation for ordered response categories." *Psychometrika*, 43(4): 561-573.

Habibzadeh F., Habibzadeh P., and Yadollahie M. (2016). "On determining the most appropriate test cut-off value: the case of tests with continuous results." *Biochemia Medica*. 26(3):297–307.

Liu X. (2012). "Classification accuracy and cut point selection." Stat Med. 31:2676–286.

Lopez-Raton M., Rodriguez-Alvarez M.X, Cadarso-Suarez C. and Gude-Sampedro F. (2014). "OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests." *Journal of Statistical Software* 61(8):1-36.

Magis D. and Raiche G. (2012). "Random Generation of Response Patterns under Computerized Adaptive Testing with the R Package catR." *Journal of Statistical Software*, 48(8), 1-31. doi:10.18637/jss.v048.i08

Miller R. and Siegmund D. (1982). "Maximally selected Chi square statistics." *Biometrics*, 38:1011–1016.

Pepe MS. (2003). "The Statistical Evaluation of Medical Tests for Classification and Prediction." *Oxford University Press: New York*.

Perkins N.J. and Schisterman E.F. (2006). "The inconsistency of optimal cutpoints obtained using two criteria based on the receiver operating characteristic curve." *Am. J. Epidemiol.* 163:670–675.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rota M. and Antolini L. (2014). "Finding the optimal cut-point for Gaussian and Gamma distributed biomarkers." *Computational Statistics & Data Analysis*,69:1–14.

Youden WJ. (1950). "Index for rating diagnostic tests." *Cancer*, 3:32–35.

Zhou X.H., Obuchowski N.A., and McClish D.K. (2011). "*Statistical Methods in Diagnostic Medicine*." Wiley: New York