ARTICLE IN PRESS

Journal of the Korean Statistical Society ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Nonparametric testing under crossover design for ordered categorical response

Joydeep Basu [a], Uttam Bandyopadhyay [b,*], Ganesh Dutta [a]

[a] *Basanti Devi College, India*
[b] *University of Calcutta, India*

ARTICLE INFO

Article history:
Received 16 July 2016
Accepted 27 March 2017
Available online xxxx

AMS 2000 subject classifications:
primary 62G86
secondary 62K86

Keywords:
Ordinal scale
Ridit
Crossover design
Score test
H-decomposition

ABSTRACT

This paper uses ridit introduced by Bross (1958) to develop a nonparametric procedure for testing the difference between two treatment effects with ordered categorical outcomes under two period crossover design in which subjects receive same or different treatment during the period of study. The performance of this procedure is compared numerically with that of a procedure obtained from Armitage (1955) in terms of type I error rate and power. The numerical study shows the ridit approach to tend to relatively more efficient in terms of power as the number of categories increases. The procedures are illustrated using the data from Matthews (1989) on crossover trial with two hypertensive agents.

1. Introduction

In this work we formulate an asymptotic procedure for the comparison between two treatments with ordinal scale responses under a $(2 \times 2 \times 4)$ crossover design (Balaam, 1968; Chow & Liu, 2008), in which two treatments, A and B, are used in two periods in a specified sequence from {AA, AB, BA, BB}. The ordinal scale is assumed to have L qualitatively ordered categories represented on the scale $\{1, 2, \ldots, L\}$ and the responses are recorded according to their level of satisfaction. Let $\widetilde{Y}_1$ and $\widetilde{Y}_2$ be two independent responses on a continuous ordinal scale corresponding to treatments A and B, respectively. Then the quantity $\widetilde{\theta}$=P($\widetilde{Y}_2 < \widetilde{Y}_1$) represents in terms of probability that the treatment under A is better for the subject compared to treatment under B. In ridit analysis, a proxy for $\widetilde{\theta}$ can be obtained by the functional

$$\theta = \mathrm{P}(Y_2 < Y_1) + \frac{1}{2}\mathrm{P}(Y_1 = Y_2), \tag{1.1}$$

called ridit reliability functional (Bross, 1958; Beder & Heim, 1990), where Y_1 and Y_2 are two independent discrete random variables on $\{1, 2, \ldots, L\}$ with P($Y_m = j$) = P($\widetilde{Y}_m \varepsilon$ category j), j = 1, 2,…, L; $m = 1, 2$. Here we interpret θ as the effect due to treatment A relative to that due to treatment B. In the subsequent development we find three different θ depending upon the response probabilities in two periods. The importance of the present work using such θ is that it ignores qualitative assessment of the categories and works with their natural ordering. Moreover, it helps to construct tests in accordance with the structure of null and alternative hypotheses. In this connection, it would be worthwhile to mention that a subject under

* Corresponding author.
E-mail addresses: joydeepstat@gmail.com (J. Basu), ubandyopadhyay08@gmail.com (U. Bandyopadhyay), duttaganesh78@gmail.com (G. Dutta).

http://dx.doi.org/10.1016/j.jkss.2017.03.003
1226-3192/

A (or B) will be at a higher satisfaction level compared to the subject under B (or A) if $P(Y_2 < Y_1) > (\text{or} <) P(Y_2 > Y_1)$, which, on simplification, becomes equivalent to $\theta > (\text{or} <)\frac{1}{2}$. On the other hand, subjects under A will be neither at a higher nor at a lower satisfaction level compared to the subjects under B if $P(Y_2 < Y_1) = P(Y_2 > Y_1)$, or equivalently, $\theta = \frac{1}{2}$, and hence A and B would be considered as equivalent. The use of θ for comparing two treatments can be found, among others, in the works of Bandyopadhyay and Biswas (2001), Bandyopadhyay and De (2007, 2009), Bandyopadhyay and Mukherjee (2015) and Brunner and Munzel (2000).

In the available literature, the inferential procedures for comparing two treatments with multiple, $L > 2$, ordered categories correspond to parallel arm design in which the data can be arranged in a $2 \times k$ contingency table. These procedures include the statistical analysis based on the large sample χ^2-test and various exact, conditional and approximate tests based on Wilcoxon mid-rank statistic. See, for example, Shan and Ma (2016). However, the χ^2-test fails to incorporate the natural ordering of the categories and the large sample mid-rank test may not be appropriate in clinical trials, as mentioned in Shan and Ma (2016), for its failure to control the rates of type I error from small size to medium size sample. On the other hand, there is an extensive work on both parametric, which includes normal and binary responses, and the nonparametric approaches for analysis of crossover data using two or more treatments. See Bandyopadhyay, Biswas and Bhattacharya (2011), Bandyopadhyay and Chatterjee (2015), Jaki, Pallmann and Wolfsegger (2013), Liang, Li, Wang and Carriere (2014) and Putt and Chinchilli (2004), among others. However, except the work of Tudor and Koch (1994), no other updated reference has yet been found for comparing two ordered multinomials $(2 \times 2 \times 4)$ crossover design. Let us address this in the framework of hypothesis testing through three different θ in two periods. In the crossover design considered here, a subject remains on the same treatment or switched to another treatment during the period of study. As a result, proper utilization of intra subject comparison reduces the bias due to different subjects and also results larger power than that obtained from inter subject comparison while testing for the difference between treatment effects. On the other hand, intra subject comparison is very much affected by the presence of possible carryover effects. As inter subject comparison has no such difficulties, such measures can be exclusively considered in the present analysis. In this connection, the works of Elswick and Uthoff (1989), Putt and Chinchilli (2004) and Wallenstein and Fisher (1977), among others have been found to be based on intra subject evaluation, whereas the works of Jung and Koch (1999), Kawaguchi and Koch (2010) and Kawaguchi, Koch and Ramaswamy (2009), among others deal with inter subject evaluation.

The layout of the paper is as follows. In Section 2, we frame our null hypothesis for testing equivalence of two treatments against two sided general alternative in $(2 \times 2 \times 4)$ crossover design. This section includes, along with some asymptotic results, the formulation of the proposed test and its competitor based on score statistics. Next, in Section 3, we consider ordered alternative and provide naive tests corresponding to the procedures in Section 2. Further, we separately deal with the particular case $L = 2$ in Section 4. In Section 5, we carry out numerical computation via simulation to compare the proposed procedure with the alternative procedure through Type I error rate and power of the respective tests. In Section 6, we provide a data study to justify the use of the procedures. A brief discussion concludes the article in Section 7.

2. Tests for equivalence against two sided alternative

In this section, we provide asymptotic procedures based on data under crossover design to formulate the proposed test from ridit statistics and its nonparametric competitor from score statistics.

2.1. Crossover design

Let us consider two treatments, A and B. The number of subjects to be examined is N, a pre-fixed positive integer. Each subject is treated twice in one of four treatment sequences: AA, AB, BA, BB. Without loss of generality, we allocate treatment A on the first N_A subjects and treatment B on the next N_B subjects in the first period, where $N_A + N_B = N$. In the second period, we split N_A subjects into two parts, in which N_{Ak} subjects correspond to treatment k in period 2 under the treatment sequence Ak, $k = $ A, B, where $N_{AA} + N_{AB} = N_A$. Similarly, N_{Bk} denotes the number of subjects that are examined under the treatment sequence Bk, $k = $ A, B, where $N_{BA} + N_{BB} = N_B$. Here N_k and $N_{kk'}$ are all pre-fixed positive integers, where $k,\ k' = $ A, B.

Crossover designs employ a simple randomization scheme that equally allocates subjects to each treatment group or treatment sequence. Equal randomization often allows statistically efficient and powerful comparisons of the treatment effects. Moreover, it has been shown in Stufken (1996) that a uniform strongly balanced crossover design for independently and identically distributed continuous responses is universally optimal for the estimation of direct treatment effects and residual (carryover) treatment effects, respectively. Therefore, in the present study, we consider equal allocation for computation purpose, and hence we take $N_{kk'} = n$, say, $k, k' = $ A, B.

2.2. Ridit test

Let X_k be a discrete random variable on $\{1,\ 2,\ \ldots,\ L\}$ corresponding to the response in the first period when kth treatment is applied, where $k = $ A, B. Then we define the following function for responses in period 1:

$$\varphi(X_A, X_B) = \begin{cases} 1; & X_A > X_B \\ \frac{1}{2}; & X_A = X_B \\ 0; & \text{otherwise.} \end{cases} \tag{2.1}$$

Please cite this article in press as: Basu, J., et al., Nonparametric testing under crossover design for ordered categorical response. Journal of the Korean Statistical Society (2017), http://dx.doi.org/10.1016/j.jkss.2017.03.003.