

Some New Results on Permutation Tests for Ordinal Categorical Variables under Restricted Alternatives

Alcuni Nuovi Sviluppi sui Test di Permutazione per Variabili Categoricali Ordinali e Alternative Ristrette

Rosa Arboretti Giancristofaro

Stefano Bonnini

Dipartimento di Scienze Statistiche, Università di Padova, via Cesare Battisti 241,
35121 Padova, e-mail: bonnini@stat.unipd.it

Riassunto: Nel presente contributo sono discussi alcuni test di permutazione per il problema del confronto tra due campioni nel caso di variabili categoriali ordinali, nell'ambito della metodologia della combinazione non parametrica di test di permutazione dipendenti. In particolare viene illustrato un nuovo test di permutazione basato sulla combinazione di un numero finito di momenti campionari. A conclusione del lavoro sono riportati i risultati di uno studio di simulazione in cui le soluzioni di permutazione sono confrontate con il test di Wilcoxon con correzione per i ties.

Keywords: ordered categorical data, restricted alternatives

1. Introduction

Stochastic dominance problems in testing for ordered categorical variables are of specific interest because they are frequently encountered in practice and present peculiar difficulties, especially within the framework of likelihood ratio tests (Agresti, 1992; Liu and Agresti, 2005; Silvapulle and Sen, 2005). Until now, the literature essentially considers the univariate case and several solutions have been proposed to cope with, most of which are based on the restricted maximum likelihood ratio test. These solutions are generally criticized, because the degree of accuracy of their asymptotic null and alternative distributions is difficult to assess and to characterize. Pesarin (2001), by working within the nonparametric combination of dependent permutation tests (NPC), proposes exact solutions to that kind of problems. In the present paper we will review some of these and discuss a new exact solution based on joint analysis of a finite set of sampling moments of ranks, or general scores, assigned to ordered classes and processed by the NPC method. Some Monte Carlo simulations for power comparisons are also presented.

2. Some Permutation Solutions

Let us assume that the support of a univariate non-degenerate ordered categorical variable X is partitioned into a finite number $K \geq 2$ of ordered classes $(A_1, A_2, \dots, A_K)$, and that the data are classified according to two levels of a treatment, giving rise to a

typical two-sample problem. Thus, given two independent samples of respectively $n_j > 2, j = 1, 2$, independent and identically distributed (iid) observations, we want to test:

$$H_0 : \left\{ X_1 \stackrel{d}{=} X_2 \right\} = \{ F_1(A_k) = F_2(A_k), k = 1, \dots, K-1 \},$$

against $H_1 : \left\{ X_1 \stackrel{d}{>} X_2 \right\} = \{ F_1(A_k) \leq F_2(A_k), k = 1, \dots, K-1 \}$, where at least one

inequality is strict and the function $F_j(A_k) = \Pr\{X_j \leq A_k\}$ plays the role of cumulative distribution function for $X_j, j = 1, 2$. In this context, by assuming that no reverse inequality such as $F_1(A_k) > F_2(A_k), k = 1, \dots, K-1$, is possible, the alternative can also be written in the form $H_1 : \left\{ \bigcup_{k=1}^{K-1} [F_1(A_k) < F_2(A_k)] \right\}$. In the context of ordered categorical variables, observed data are generally organized in a $2 \times K$ contingency table such as $\{f_{jk} = \sum_{i \leq n_j} \mathbf{I}(X_{ji} \in A_k), k = 1, \dots, K, j = 1, 2\}$, where $\mathbf{I}(\bullet) = 1$ if event $(\bullet)$ occurs and 0 otherwise. Symbols $N_{jk} = \sum_{s \leq k} f_{js}, n_j = N_{jK}$ and $f_{\bullet k} = f_{1k} + f_{2k}$, indicate cumulative and marginal frequencies, respectively. Permutation analysis is much easier if, in place of usual contingency tables, data are unit-by-unit represented by listing the $n = n_1 + n_2$ individual records. In the $2 \times K$ design, data set is represented by $\mathbf{X} = \{X(i), i = 1, \dots, n; n_1, n_2\}$. If $(u_1^*, \dots, u_n^*)$ indicates a permutation of individual units $(1, \dots, n)$, then $\mathbf{X}^* = \{X(u_i^*), i = 1, \dots, n; n_1, n_2\}$ indicates the corresponding permutation of data set $\mathbf{X}$, whose associated permuted table is $\{f_{jk}^* = \sum_{i \leq n_j} \mathbf{I}(X_{ji}^* \in A_k), k = 1, \dots, K, j = 1, 2\}$, where $X_{1i}^* = X(u_i^*)$ for $i \leq n_1$ and $X_{2i}^* = X(u_i^*)$ for $n_1 < i \leq n$. In this framework, as a solution to the two-sample one dimensional testing problem we may consider the permutation test statistic:

$$T_{AD}(\mathbf{X}^*) = T_{AD}^* = \sum_{k=1}^{K-1} \left(\hat{F}_{2k}^* - \hat{F}_{1k}^* \right) \left[\bar{F}_{\bullet k} (1 - \bar{F}_{\bullet k}) \frac{4n_1}{n_2(n-1)} \right]^{-1/2},$$

where $\hat{F}_{jk}^* = N_{jk}^* / n_j, j = 1, 2, \bar{F}_{\bullet k} = N_{\bullet k} / n$ are permutation and marginal empirical distribution functions (EDFs) respectively, and $N_{1k}^*, N_{2k}^*, k = 1, \dots, K-1$ are permutation cumulative frequencies. Note that large values for T_{AD} are significant. Statistic T_{AD} essentially compares two EDFs and corresponds to the discrete version of a statistic following Cramér-von Mises' two-sample goodness-of-fit statistic for stochastic dominance alternatives, adjusted according to Anderson-Darling. The p -value is defined as $\lambda_{AD} = \Pr\{T_{AD}^* \geq T_{AD}^o | \mathbf{X}\}$, where $T_{AD}^o = T_{AD}(X)$ represents the observed value. Thus, according to the general testing rule, if $\lambda_{AD} \leq \alpha$ the null hypothesis is rejected at significance level $\alpha > 0$. Let us observe that the k th summand ($k = 1, \dots, K-1$)

in T_{AD} may be seen as a partial test for the sub-hypotheses $H_{0k} : \{F_k = G_k\}$ and $H_{1k} : \{F_k < G_k\}$. Thus the global hypotheses may equivalently be written as $H_0 : \bigcap_k H_{0k}$ and $H_1 : \bigcup_k H_{1k}$, where a suitable break down into a set of partial sub-hypotheses is emphasized. Hence, in order to obtain an overall solution one way is to properly combine all related partial results. Of course, these partial tests and associated p -values are dependent in a way that in general is extremely difficult to take into account explicitly, consequently, when considering their combination we shall take account nonparametrically of their underlying dependence relations; hence we shall work within the NPC approach. Theory and methods for this kind of solutions are fully discussed in Pesarin (2001). In the NPC approach we need combining the p -values λ_k associated with partial tests by a non-degenerate and measurable combining function ψ (for example, Fisher's: $T_F'' = -\sum_k \log(\lambda_k)$, Tippett's: $T_T'' = \max_k (1 - \lambda_k)$, etc.). As a frequently used combining function there is the so-called direct combination consisting in a function of partial test statistics instead of related p -values (for example $T_D'' = \sum_k T_k$, where T_k denotes the partial test). Thus, T_{AD} may be seen as a direct nonparametric combination and so it enjoys all NPC properties.

With reference to the $2 \times K$ design, one more application of NPC methodology is in terms of joint analysis of tests on sampling moments. To this end, let us assign ranks W to ordered classes, that is let us transform A_k into $W_k = k$, and consider the rule: two discrete distributions defined on the same support, with a finite number K of distinct real values, are equal if and only if their first $K-1$ moments are equal, because their characteristic functions, as well as their probability generating functions, depend only on these few moments. Consequently, we are allowed to write global hypotheses as

$$H_0 : \left\{ X_1 \stackrel{d}{=} X_2 \right\} = \left\{ \bigcap_{r=1}^{K-1} \mathbf{E}(W_1^r) = \mathbf{E}(W_2^r) \right\}, \quad H_1 : \left\{ X_1 \stackrel{d}{>} X_2 \right\} = \left\{ \bigcup_{r=1}^{K-1} \mathbf{E}(W_1^r) > \mathbf{E}(W_2^r) \right\}.$$

Let us now consider for each $r, r = 1, \dots, K-1$, the permutation partial test statistic based on comparison of the r th sampling moments or on one of its permutationally equivalent statistics, such as $T_{W_r}^* = \sum_{k \leq K} k^r f_{1k}^* / n_1$. We note that all these partial tests are exact, unbiased, and consistent, then an NPC test associated with the combining function ψ is $T_{W\psi}'' = \psi(\lambda_{W1}, \dots, \lambda_{WK-1})$; and therefore, for any ψ , combined tests $T_{W\psi}''$ are at least exact, unbiased, and consistent.

3. Results of a comparative simulation study and concluding remarks

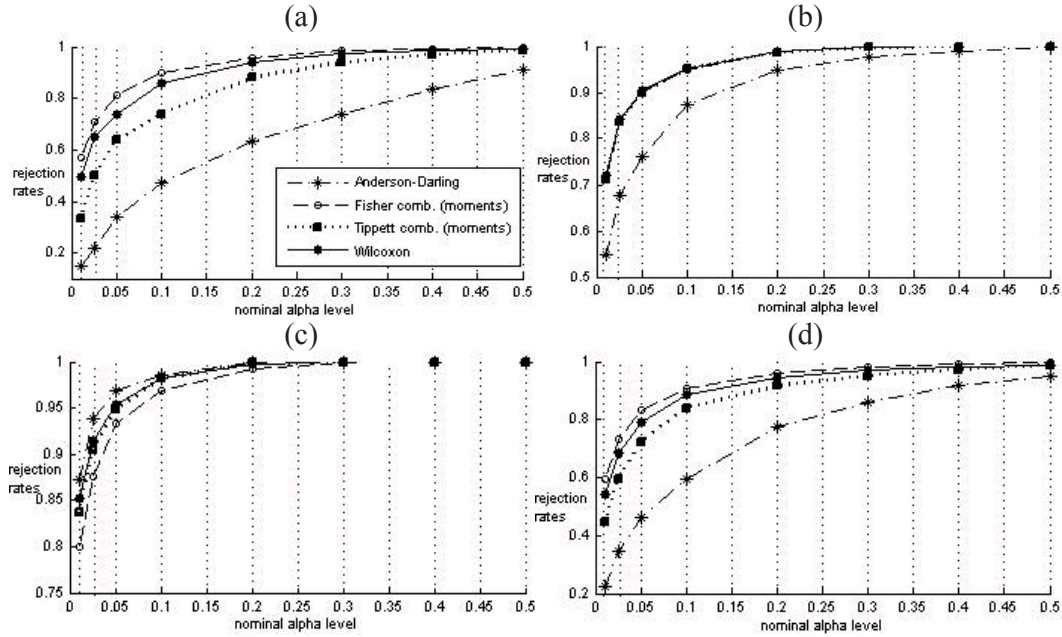
The behaviour of the following tests has been examined: the Wilcoxon rank sum test W with ties correction and large sample approximation; the Anderson-Darling permutation test T_{AD} ; the permutation test T_{WF}'' based on Fisher's combination of sampling moments; the permutation test T_{WT}'' based on Tippett's combination of sampling moments. Data were generated from ordinal distributions with four categories. Table 1, shows the distributional settings from (a) to (d) considered for the two samples under the restricted alternative hypothesis $H_1 : \{ X_1 \stackrel{d}{>} X_2 \}$ with sample sizes of $n_1 = n_2 = 30$

(results not shown) and $n_1 = 30$, $n_2 = 20$. For each configuration we perform 1000 random data generations and permutation tests are performed with 1000 Monte Carlo permutations. Under the null hypothesis, all tests showed achieved significance levels close to nominal ones. Quite different behaviour among tests resulted under the alternative (Fig. 1). Anderson-Darling test is more powerful only in configuration reported in figure 1(c). Permutation test T_{WF}'' showed a good behaviour in all situations both for balanced and unbalanced sample sizes and, with the exception of distributional settings 1(c) and 1(b) for $n_1 = 30$, $n_2 = 20$, it is more powerful than Wilcoxon test.

Table 1: Frequency distributions (%) in each category of the two samples

Category	Configuration (a)				Configuration (b)				Configuration (c)				Configuration (d)			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Sample 1	5	10	15	70	5	10	15	70	5	10	15	70	5	10	15	70
Sample 2	5	10	55	30	5	50	15	30	45	10	15	30	5	20	45	30

Figure 1: Empirical power for $n_1 = 30$, $n_2 = 20$ for configurations (a)-(d)



References

- Agresti, A. (1992), A survey of exact inference for contingency tables. *Statistical Science*; 7, 131-177.
- Liu, I., Agresti, A. (2005), The analysis of ordered categorical data: an overview and a survey of recent developments. *Test*; 14, 1-73.
- Pesarin, F. (2001) *Multivariate Permutation Test With Application to Biostatistics*, Wiley, Chichester.
- Silvapulle, M.J., Sen, P.K. (2005). *Constrained Statistical Inference, Inequality, Order, and Shape Restrictions*. Wiley, New York.