

Finite sample consistency for non associative statistics

Livio Corain¹, Fortunato Pesarin², Luigi Salmaso¹

¹ Dipartimento di Tecnica e Gestione dei sistemi industriali, Università di Padova

² Dipartimento di Scienze Statistiche, Università di Padova

e-mail: livio.corain@unipd.it; pesarin@stat.unipd.it; salmaso@gest.unipd.it

Abstract: A quite important problem usually occurs in some multidimensional applications when sample sizes are fixed and the number of variables which are to be analyzed is much larger than sample sizes (Blair et al., 1994; Goggin, 1986). In Pesarin (2001) it is shown that, under very mild conditions, the power function of permutation tests based on associative statistics monotonically increases as the related standardized noncentrality functional increases. This is true also for multivariate situations. In particular, for any added variable the power does not decrease if this variable makes larger standardized noncentrality (*finite-sample consistency*). Results of this paper confirm and extend those presented by Blair et al. (1994) and Pesarin and Salmaso (2009).

Keywords: nonparametric combination, permutation tests.

1. Introduction and main notation

As a guide, we refer to one-sided two-sample designs for non-negative alternatives. Extensions to non-positive and/or two-sided alternatives, and multisample designs are straightforward. Let $\mathbf{X}_j = \{X_{ji}; i = 1, \dots, n_j\} \in \mathcal{X}^{n_j}$ the independent and identically distributed (i.i.d.) sample data of size n_j from $(X, \mathcal{X}, P_j \in \mathcal{P})$, $j = 1, 2$, where X is the variable of interest taking values in the sample space $\mathcal{X}$ according to the distribution P_j , i.e. any member of the nonparametric family $\mathcal{P}$. A notation for data sets with independent samples is $\mathbf{X} = \{X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}\} \in \mathcal{X}^n$, where $n = n_1 + n_2$. To denote data sets in the permutation context it is sometimes convenient to use the unit-by-unit representation: $\mathbf{X} = \{X_i; i = 1, \dots, n; n_1, n_2\}$, where it is intended that first n_1 data in the list belong to first sample and the rest to the second. In practice, denoting by $(u_1^*, \dots, u_n^*)$ a permutation of $(1, \dots, n)$, $\mathbf{X}^* = \{X_i^* = X(u_i^*), i = 1, \dots, n; n_1, n_2\}$ is the related permutation of $\mathbf{X}$, so that $\mathbf{X}_1^* = \{X_{1i}^* = X(u_i^*), i = 1, \dots, n_1\}$ and $\mathbf{X}_2^* = \{X_{2i}^* = X(u_i^*), i = n_1 + 1, \dots, n_2\}$ are the two permuted samples, respectively.

Here we discuss testing problems for stochastic dominance alternatives as are generated by symbolic treatments with non-negative random shift effects Δ . In particular, the alternative assumes that treatments produce effects Δ_1 and Δ_2 , respectively, and that $\Delta_1 >^d \Delta_2$, where $>^d$ stands for distribution (or stochastic) dominance. Thus, the hypotheses are

$$H_0: \{X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} X\} \equiv \{P_1 = P_2\} \text{ vs. } H_1: \{(X_1 + \Delta_1) >^d (X_2 + \Delta_2)\}.$$

Note that under H_0 data of two samples are exchangeable, in accordance with the notion that subjects are randomized to treatments. Without loss of generality, we assume that effects in H_1 are such that $\Delta_1 = \Delta >^d 0$ and $\Pr\{\Delta_2 = 0\} = 1$. Condition $\Delta_2 = 0$ agrees with the notion that an active treatment is only assigned to subjects of first sample and a placebo

to those of the second. In this situation, since effects Δ may depend on null responses X_1 , stochastic dominance $(X_1 + \Delta) \succ^d X_2 = X$ is compatible with non-homoscedasticities in the alternative. Thus, the null hypothesis may also be written as $H_0: \{\Delta =^d 0\}$. Sometimes, to emphasize the role of effects we use $\mathbf{X}(\Delta) = \{X_{11} + \Delta_1, \dots, X_{1n_1} + \Delta_{n_1}, X_{21}, \dots, X_{2n_2}\}$ to denote data sets, and so $\mathbf{X}(0)$ denotes data in H_0 . Without loss of generality we can model the data set as $\mathbf{X}(\Delta) = \{\mathbf{Z}_1 + \Delta, \mathbf{Z}_1\}$ where $\Delta = \{\Delta_{11}, \dots, \Delta_{1n_1}\}$ and where Z have the role of random deviates the distribution P_Z of which is generally unknown. In the context of this work, it is also worth noting that observed variable X , random deviates Z , sample space $\mathcal{X}$, and random effect Δ are V -dimensional, with $V \geq 1$.

Permutation tests are known to be conditional methods of inference, where the conditioning is made with respect to a set of sufficient statistics for P in H_0 (Pesarin, 2001; Randles and Wolfe, 1979), the related conditional reference space is denoted by $\mathcal{X}^n_{/X}$. Sufficiency of $\mathcal{X}^n_{/X}$ in H_0 implies that the null conditional probability of every measurable event $A \subset \mathcal{X}^n$, given $\mathcal{X}^n_{/X}$, is independent of P ; that is, with clear meaning of the symbols, $\Pr\{\mathbf{X}^* \in A; P | \mathcal{X}^n_{/X}\} = \Pr\{\mathbf{X}^* \in A | \mathcal{X}^n_{/X}\}$. In this paper we consider test statistics based on comparison of sampling indicators like $T^*(\Delta) = S(\mathbf{X}_1^*(\Delta)) - S(\mathbf{X}_2^*(\Delta))$, where $S(\mathbf{X}_j^*(\Delta)): \mathcal{X}^{n_j} \rightarrow \mathcal{R}^1$ is any function invariant with respect to rearrangements of entry arguments: $S(X_1, \dots, X_k) = S(X_{u_1}, \dots, X_{u_k})$ in which $u_1, \dots, u_k$ is any rearrangement (permutation) of $1, \dots, k$. These kind of statistics include associative forms like

$$T^*(\Delta) = 1/n_1 \sum_i \varphi[X_{1i}^*(\Delta)] - 1/n_2 \sum_i \varphi[X_{2i}^*(\Delta)],$$

where φ is any non-degenerate measurable non-decreasing function of the data and so T^* corresponds to the comparison of sampling φ -means: $T^* = \bar{\varphi}_1^* - \bar{\varphi}_2^*$ say. Moreover, they also include non-associative statistics like $T^*(\Delta) = \varphi[\tilde{\mathbf{X}}_1^*(\Delta)] - \varphi[\tilde{\mathbf{X}}_2^*(\Delta)]$ corresponding to the comparison of sampling φ -medians: $T_- = \tilde{\varphi}_1^* - \tilde{\varphi}_2^*$, etc.

Finite-sample consistency is based on the nonparametric combination (NPC) of partial dependent tests, one for each component or derived variable. Within the NPC procedure all partial tests are assumed to be conditional and unconditional unbiased. Proof of unbiasedness is based on point-wise comparisons of values assumed by $T^*(0)$ and $T^*(\Delta)$ in each permutation, relative to their respective observed values $T(0)$ and $T(\Delta)$.

2. Finite-sample consistency for non-associative statistics

In this section, we give a short overview of main theory behind the finite-sample consistency property for associative statistics first. Let us suppose that T is any associative test statistic for one-sided hypotheses, sample sizes (n_1, n_2) and the set of real deviates $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) \in \mathcal{X}^n$, are fixed. Without loss of generality, the data set is $\mathbf{X}(\delta) = (\mathbf{Z}_1 + \delta, \mathbf{Z}_2)$, where the $\mathbf{Z}$ are i.i.d. measurable real random deviates, whose parent distribution is $P_Z(z) = \Pr\{Z \leq z\}$ and $\delta = (\delta_1, \dots, \delta_1)$ is the n_1 -dimensional vector of non-negative fixed effects. Suppose also that effects diverge to the infinity according to whatever monotonic sequence $\{\delta_v, v \geq 1\}$, the elements of which are such that $\delta_v \leq \delta_{v'}$ for any pair $v < v'$. If those conditions are satisfied, then the permutation (conditional) rejection rate of T converges to 1 for all α -values not smaller than the minimum attainable α_a ; thus, T is conditional [i.e., for fixed $(\mathbf{Z}_1, \mathbf{Z}_2)$] and unconditional [i.e., for $\forall (\mathbf{Z}_1, \mathbf{Z}_2) \in \mathcal{X}^n$] finite-sample consistent. Furthermore, suppose that random deviates Z and effects δ are such that there exists a function $\rho(\delta) > 0$ of effects δ the limit of which

is 0 as δ goes to the infinity, T is any associative test statistic as above, and the data set is obtained by considering the transformation $\mathbf{Y}(\delta) = \rho(\delta)\mathbf{X}(\delta)$. If $\lim_{\delta \rightarrow \infty} \delta \rho(\delta) = \tilde{\delta} > 0$, and $\lim_{\delta \rightarrow \infty} \Pr\{\rho(\delta) \cdot |Z| > \varepsilon\} = 0$, $\forall \varepsilon > 0$, then the unconditional rejection rate converges to 1 for α -values not smaller than the minimum attainable α_a ; and thus T is weak unconditional finite-sample consistent (Pesarin and Salmaso, 2009). The extension of these results to random effects Δ is also shown in Pesarin and Salmaso (2009).

For instance, suppose a problem in which the V -dimensional data set is

$$\mathbf{X}(\delta) = (\delta_h + \sigma_h Z_{h1i}, i=1, \dots, n_1, \sigma_h Z_{h2i}, i=1, \dots, n_2; h=1, \dots, V),$$

where δ_h and σ_h are the fixed effect and the scale coefficient of the h th component variable, respectively, the hypotheses are $H_0: \{\mathbf{X}_1 =^d \mathbf{X}_2\} = \{\delta=0\}$ against $H_1: \{\mathbf{X}_1 >^d \mathbf{X}_2\} = \{\delta \geq 0\}$. Suppose also that the test statistic is $T_V^*(\delta) = 1/V \sum_{h=1}^V [\bar{X}_{h1}^*(\delta_h) - \bar{X}_{h2}^*(\delta_h)]/S_h$, where $\bar{X}_{hj}^*(\delta_h) = \sum_i [X_{hji}^*(\delta_h)/n_j]$ is the permutation mean of j th sample and S_h a permutation invariant statistic for the h th scale coefficient σ_h , i.e. a function $S[X_{hji}(\delta_h), i=1, \dots, n_j, j=1, 2]$ of pooled data such as for instance $S_h = \text{Md}[|X_{hji} - \tilde{X}_h|, i=1, \dots, n_j, j=1, 2]$ the median of absolute deviations from the median specific to the h th variable. It can be proved that a sufficient condition for finite-sample consistency of $T_V^*(\delta)$ is that all Z_h have finite first moment. Thus, when some of the multivariate components do not possess finite mean value, a test based on comparisons of sampling means is not finite-sample consistent. It is worth noting that $T_V^*(\delta)$ is the direct combination (Pesarin, 2001) of V partial tests $\bar{X}_{h1}^*(\delta_h) - \bar{X}_{h2}^*(\delta_h)$. In order to extend finite-sample consistency to non-associative statistics, let us briefly introduce the notion of conditional (permutation) unbiasedness for any kind of statistics $T^*(\Delta) = S(\mathbf{X}_1^*(\Delta)) - S(\mathbf{X}_2^*(\Delta))$. To this end and with clear meaning of the symbols, let us observe that:

- $T^o(0) = S(\mathbf{Z}_1) - S(\mathbf{Z}_2)$, i.e. the null observed value of statistic T .
- $T^o(\Delta) = S(\mathbf{Z}_1 + \Delta_1) - S(\mathbf{Z}_2) = S(\mathbf{Z}_1) + D_S(\mathbf{Z}_1, \Delta_1) - S(\mathbf{Z}_2) = T^o(0) + D_S(\mathbf{Z}_1, \Delta_1)$, where $D_S(\mathbf{Z}_1, \Delta_1) \geq 0$.
- $T^*(0) = S(\mathbf{Z}_1^*) - S(\mathbf{Z}_2^*)$, i.e. the value of T in the permutation $\mathbf{u}^* = u_1^*, \dots, u_n^*$.
- $T^*(\Delta) = S(\mathbf{Z}_1^* + \Delta_1^*) - S(\mathbf{Z}_2^* + \Delta_2^*) = T^*(0) + D_S(\mathbf{Z}_1^*, \Delta_1^*) - D_S(\mathbf{Z}_2^*, \Delta_2^*)$.
- $D_S(\mathbf{Z}_2^*, \Delta_2^*) \geq D_S(\mathbf{Z}_2^*, 0) = 0 = D_S(\mathbf{Z}_2, 0)$, because effects Δ_{2i}^* coming from first group are non-negative.
- $D_S(\mathbf{Z}_1^*, \Delta_1^*) \leq D_S(\mathbf{Z}_1^*, \Delta_1)$ point-wise, because in $D_S(\mathbf{Z}_1^*, \Delta_1^*)$ there are non-negative effects assigned to units coming from group 2; e.g., suppose $n_1=3$, $n_2=3$, and $\mathbf{u}^* = (3, 5, 4, 1, 2, 6)$, then $(\mathbf{Z}_1^*, \Delta_1^*) = [(Z_{13}, \Delta_{13}), (Z_{22}, 0), (Z_{21}, 0)]$, and so $(\mathbf{Z}_1^*, \Delta_1) = [(Z_{13}, \Delta_{13}), (Z_{22}, \Delta_{11}), (Z_{21}, \Delta_{12})]$,

or

$$(\mathbf{Z}_1^*, \Delta_1) = [(Z_{13}, \Delta_{13}), (Z_{22}, \Delta_{12}), (Z_{21}, \Delta_{11})];$$

it is to be emphasized that $X(u_i^*) = Z(u_i^*) + \Delta(u_i^*)$ if $u_i^* \leq n_1$, that is units coming from first group maintain their effects, whereas the rest of effects are randomly assigned to units coming from second group.

- $D_S(\mathbf{Z}_1^*, \Delta_1) =^d D_S(\mathbf{Z}_1, \Delta_1)$, because $\Pr\{\mathbf{Z}_1^* | X_{/X(0)}\} = \Pr\{\mathbf{Z}_1 | X_{/X(0)}\}$ (see Pesarin, 2001).

Thus, $D_S(\mathbf{Z}_1^*, \Delta_1^*) - D_S(\mathbf{Z}_2^*, \Delta_2^*) \leq D_S(\mathbf{Z}_1, \Delta_1)$ in permutation distribution and so

$$\begin{aligned} \lambda_T(\mathbf{X}(\Delta)) &= \Pr\{T(\mathbf{X}^*(\Delta)) \geq T(\mathbf{X}(\Delta)) | X_{/X(\Delta)}\} \\ &= \Pr\{T^*(0) + D_S(\mathbf{Z}_1^*, \Delta_1^*) - D_S(\mathbf{Z}_2^*, \Delta_2^*) - D_S(\mathbf{Z}_1, \Delta_1) \geq T^o(0) | X_{/X(0)}\} \\ &\leq \Pr\{T^*(0) \geq T^o(0) | X_{/X(0)}\} = \lambda_T(\mathbf{X}(0)), \end{aligned}$$

which establishes the dominance in permutation distribution of $\lambda_T(\mathbf{X}(\Delta))$ with respect to $\lambda_T(\mathbf{X}(0))$, uniformly for all data sets $\mathbf{X} \in \mathcal{X}^n$, for all underlying distributions P , and for all associative and non-associative statistics $T = S(\mathbf{X}_1) - S(\mathbf{X}_2)$. If effects were fixed, $\Delta = \mathbf{p}$

say, then $D_S(\mathbf{Z}_1^*, \delta_1^*) - D_S(\mathbf{Z}_2^*, \delta_2^*) \leq D_S(\mathbf{Z}_1, \delta_1)$ is point-wise, i.e. somewhat stronger than in permutation distribution. For instance, these results allows us to prove the following

Theorem. Suppose that in a two-sample problem there are $V \geq 1$ non homoscedastic variables $X = (X_1, \dots, X_V)$, the observed data set is $\mathbf{X}(\delta) = (\delta_h + \sigma_h Z_{h1i}, i=1, \dots, n_1, \sigma_h Z_{h2i}, i=1, \dots, n_2, h=1, \dots, V)$ and the hypotheses are $H_0: \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2\} = \{\delta = 0\}$ against $H_1: \{\mathbf{X}_1 \stackrel{d}{>} \mathbf{X}_2\} = \{\delta \geq 0\}$, where $\delta = (\delta_1, \dots, \delta_V)'$. For the testing purpose consider the statistic

$$T_{Md}^*(\delta) = 1/n_1 \sum_{i \leq n_1} \tilde{Y}_{1i}^*(\delta) - 1/n_2 \sum_{i \leq n_2} \tilde{Y}_{2i}^*(\delta),$$

where $\tilde{Y}_{ji}^*(\delta) = \text{Md}[X_{hji}^*(\delta)/S_{Mh}, h=1, \dots, V], i=1, \dots, n_j, j=1, 2$, is the median of V scale-free component variables specific to i th subject, and $S_{Mh} = \text{MAD}_h = \text{Md}[|X_{hji} - \tilde{X}_h|, i=1, \dots, n_j, j=1, 2]$ is the median of absolute deviations from the median specific to the variable X_h .

In this setting, the test based on $T_{Md}^*(\delta)$ is conditional and unconditional finite sample consistent as far as V diverges and $\text{Md}(Y_i(\delta)) > 0$ without requiring existence of any positive moment for V variables.

3. Conclusions

In this work we extend the so-called *finite-sample consistency* to non-associative statistics and correlated data. It is worth noting that finite-sample consistency is an important property which should be taken into consideration by the experimenter when defining the design of many observational or experimental studies. In observational studies, stratification for possible confounding factors may be adopted in order to 'create' the suitable conditions for within strata exchangeability to hold. Of course, it has to be underlined that only informative variables allow the experimenter for a gain in power of the combination procedure. An exhaustive simulation study on this topic is available at: <http://www.gest.unipd.it/~salmaso>.

Acknowledgement: The research has been supported by the University of Padova CPDA088513 (Coordinator: Prof. Luigi Salmaso).

References

- Blair, R. C., Higgins, J. J, Karniski, W., Kromrey, J. D. (1994) A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. *Multivariate Behavioral Research*, 29: 141-163.
- Goggin, M. L. (1986) The "Too Few Cases/Too Many Variables" Problem in Implementation Research. *The Western Political Quarterly*, 39: 328-347.
- Pesarin, F. (2001). *Multivariate Permutation Tests With Applications in Biostatistics*, Wiley, Chichester.
- Pesarin F., Salmaso, L. (2009) Finite-sample consistency of combination-based permutation tests with application to repeated measures designs. *Journal of Nonparametric Statistics* (to appear).
- Randles, R. H., Wolfe, D. A. (1979) *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York.