Testing for equivalence and non-inferiority: IU and UI tests within a permutation approach

Fortunato Pesarin* Luigi Salmaso† Eleonora Carrozzo† Rosa Arboretti‡

Abstract

In many research areas such as clinical trials, bioequivalence or pharmaceutical experiments, there is often the need to deal with the problem of testing the equivalence of two treatments. There are mainly two approaches with which to address the problem, the choice of which depends on the priority of the researcher who perform the analysis. The *intersection-union* principle (IU principle) considers as null hypothesis that the effect of a new treatment lies outside a given interval around that of the comparative treatment, and as alternative hypothesis that this effect lies within that interval. Alternatively, the *union-intersection* principle (UI principle) considers as null hypothesis that the effect of a new treatment lies within a given interval around that of the comparative treatment, and as alternative hypothesis that this effect lies outside that interval. Thus, given a fixed α , the researcher has to decide if it is preferable to retain with a probability converging to one an equivalence between treatments (leading to the IU approach), or a non-equivalence between treatments (leading to the UI approach). In the literature, the IU approach seems to be the only one followed, apparently without real motivations.

The goal of this paper is at first to present two practical solutions for the two approaches, working in a nonparametric setting within the permutation framework. Two algorithms respectively for IU and UI test are presented. A comparison between the behavior of the two solutions is also discussed using a simulation study.

Key Words: intersection-union principle, union-intersection principle, multi-aspect testing, non-parametric combination, permutation tests, testing for equivalence

1. Introduction

Most of the literature about testing for equivalence of two treatments, faces the problem within the framework of the *intersection-union* tests (see e.g. Berger (1982); Berger and Hsu (1996); D'Agostino *et al.* (2003); Hung and Wang (2009); Julious (2010); Laster and and Johnson (2003); Liu *et al.* (2002); Metha *et al.* (1984); Wellek (2010); Zhong *et al.* (2012)). Following the IU principle, the alternative hypothesis states that the effect of a new treatment (typically a drug) lies within a given interval around that of the comparative treatment, whereas the null hypothesis states that this effect lies outside that interval. More formally, suppose to have one endpoint variable X and a two-sample design, and we want to assess the (substantial) *equivalence* of a new treatment *A* to a comparative treatment *B*. Let δ_A be the effect of *A*, and let δ_B that of the treatment *B*, we want to test for the hypothesis $H : [(\delta_A \leq \delta_B - \varepsilon_I)$ OR $(\delta_A \geq \delta_B + \varepsilon_S)]$ against $K : (\delta_B - \varepsilon_I < \delta_A < \delta_B + \varepsilon_S)$, where $\varepsilon_I > 0$ and $\varepsilon_S > 0$ are the non-inferior and the non-superior limits for the difference $\delta_A - \delta_B$ of two effects. Limits which are suitably established by biological and/or pharmacological and/or clinical and/or technical and/or regulatory considerations.

Let us assume to break down the global hypothesis into the following partial sub-hypotheses: $H_I : \delta_A \leq \delta_B - \varepsilon_I \; vs \; K_I : \delta_A > \delta_B - \varepsilon_I$ and $H_S : \delta_A \geq \delta_B + \varepsilon_S$

*Department of Statistical Sciences, University of Padova, Italy. E-mail: pesarin@stat.unipd.it

†Department of Management and Engineering, University of Padova, Italy. E-mail: luigi.salmaso@unipd.it, carrozzo@gest.unipd.it

‡Department of Land, Environment, Agriculture and Forestry, University of Padova, Italy. E-mail: rosa.arboretti@unipd.it

$vs\ K_S : \delta_A < \delta_B + \varepsilon_S$. It is easy to observe that the global null hypothesis H is true when only one between the sub-null hypotheses H_I and H_S is true, whereas the global alternative K is true when both the sub-alternatives K_I and K_S are jointly true. Thus the hypotheses in the IU test states that $H = H_I \bigcup H_S$ and $K = K_I \bigcap K_S$.

Conversely following the UI principle, in accordance with Roy's theory (Roy, 1953; Sen, 2007), the alternative hypothesis states that the effect of a new treatment lies outside a given interval around that of the comparative treatment, whereas the null hypothesis states that this effect lies within that interval. The related hypotesis testing can be written as $\widetilde{H} : (\delta_B - \varepsilon_I \leq \delta_A \leq \delta_B + \varepsilon_S)$ against $\widetilde{K} : [(\delta_A < \delta_B - \varepsilon_I)$ OR $(\delta_A > \delta_B + \varepsilon_S)]$. As before, let us consider the related partial sub-hypotheses: $\widetilde{H}_I : \delta_A \geq \delta_B - \varepsilon_I\ vs$ $\widetilde{K}_I : \delta_A < \delta_B - \varepsilon_I$ and $\widetilde{H}_S : \delta_A \leq \delta_B + \varepsilon_S\ vs\ \widetilde{K}_S : \delta_A > \delta_B + \varepsilon_S$, and note that in this case the global null hypothesis $\widetilde{H}$ is true when both the sub-null hypotheses $\widetilde{H}_I$and $\widetilde{H}_S$ are jointly true, whereas the global alternative $\widetilde{K}$ is true if only one between the sub-alternatives $\widetilde{K}_I$and $\widetilde{K}_S$is true. The hypotheses in the UI test states that $\widetilde{H} = \widetilde{H}_I \bigcap \widetilde{H}_S$ and $\widetilde{K} = \widetilde{K}_I \bigcup \widetilde{K}_S$.

Thus, for a given fixed significance level α, the researcher who faces the problem of testing for equivalence has to decide if his focus is: the *equivalence* of the effects, i.e. when two treatments are equivalent, he wishes to detect this equivalence with probability converging to one when sample size diverges, hence follow an IU approach; or the *non-equivalence* of the effects, i.e. when two treatments are non-equivalent, he wishes to detect this non-equivalence with a probability converging to one when sample size diverges, hence follow an UI approach.

It is worth noting that the IU testing approach does not admit any solution when $\varepsilon_I = \varepsilon_S = 0$. In fact in this situation the hypothesis testing becomes $H : [(\delta_A \leq \delta_B)$ OR $(\delta_A \geq \delta_B)]$ against $K = \varnothing$. In the other hand the UI approach does not present drawbacks when $\varepsilon_I = \varepsilon_S = 0$ because the hypothesis testing becomes $\widetilde{H} : \delta_A = \delta_B$ against $\widetilde{K} : \delta_A \neq \delta_B$ which is the traditional two-sided test.

In this paper, we propose a solution for both IU and UI approaches within the permutation approach, testing separately, albeit simultaneously, the two partial test statistics, and combining them through the NonParatric Combination (NPC) procedure (Bertoluzzo *et al.*, 2013; Pesarin, 2001; Pesarin and Salmaso, 2010).

In order to introduce the proposed methods and without loss of generality, let us refer to a two-sample design, a one dimensional endpoint variable X and let us assume that a sample of n_1 IID data are from X_1 related to treatment A and, independently, n_2 IID data from X_2 related to B. Let us also suppose that the underlying variable X is common to both populations and they may differ for a shift, i.e. $X_1 = X + \delta_A$ and $X_2 = X + \delta_B$ (in this formulation it is implicitly assumed that data are homoschedastic, a condition which can be considerably weakened, see Pesarin and Salmaso (2010, 2011)). So that $\mathbf{X}_1 = (X_{11}, ..., X_{1n_1})$ are the data of the A sample and $X_2 = (X_{21}, ..., X_{2n_2})$ those of B.

2. The nonparametric IU permutation test

In this section we present the proposal for the IU approach, where the hypothesis has testing the form $H : [(\delta_A \leq \delta_B - \varepsilon_I)$ OR $(\delta_A \geq \delta_B + \varepsilon_S)]$ against $K : (\delta_B - \varepsilon_I < \delta_A < \delta_B + \varepsilon_S)$. As seen in the previous section it is possible to see the global test as splitted into two partial tests $H_I : \delta_A \leq \delta_B - \varepsilon_I\ vs\ K_I : \delta_A > \delta_B - \varepsilon_I$ and $H_S : \delta_A \geq \delta_B + \varepsilon_S\ vs\ K_S : \delta_A < \delta_B + \varepsilon_S$. The idea is to test separately, albeit simultaneously H_I vs K_I and H_S vs K_S . For testing H_I vs K_I let us consider the data X_2 of sample B modified as X_{I2}

$= X_2 - \varepsilon_I$ whereas the data of sample A are the same so that $X_{I1} = X_1$. Now we can see that the sub-alternative $K_I : \delta_A > \delta_B - \varepsilon_I \equiv X_{I1} \overset{d}{>} X_{I2}$ and $H_I : \delta_A \leq \delta_B - \varepsilon_I$, as a test for stochastic dominance between two populations. For this kind of problem a suitable test statistic may be based on the differences of the two means, i.e. $T_I = \bar{X}_{I1} - \bar{X}_{I2}$where $X_{Ij} = \sum_{i=1}^{n_j} X_{Iji}/n_j, j = 1, 2$. Similarly for testing H_S vs K_S let us consider $X_{S1} = X_1$ and data of sample B modified as $X_{S2} = X_2 + \varepsilon_S$. The hypothesis testing can be written as $K_S : \delta_A < \delta_B + \varepsilon_S \equiv X_{S1} \overset{d}{<} X_{S2}$ and $H_S : \delta_A \geq \delta_B$ and a suitable test statistic is $T_S = \bar{X}_{S2} - \bar{X}_{S1}$. Note that both test statistics T_I and T_S are significant for large values, i.e. large value are evidence agaist the respective null hypotheses. Before presenting a complete algorithm for the IU permutation test, let us introduce a modification in the NPC method, since the two partial p-values are not positively dependent. Indeed, it is worth noting that H_I true implies H_S false and vice versa.Thus the combination can be done by a nonparametric combining functions $\varphi : [0, 1]^2 \to \mathbb{R}^+$, small values of which are significant and combining functions suitable for IU testing should satisfy the following properties:

1) φ is continuous and non-decreasing in each argument, i.e. $\lambda_q < \lambda'_q$ implies $\varphi(\ldots, \lambda_q, \ldots) \leq \varphi(\ldots, \lambda'_q, \ldots)$;

2) φ must attain its infimum if all arguments attain 0;

3) $\alpha > 0$ and larger than the minimum attainable value (Pesarin and Salmaso, 2010) implies the conditional critical value $\varphi_\alpha > 0$.

Examples of combining function for the IU permutation test are:
-the max-*p* rule: $\varphi_M = \max(\lambda_I, \lambda_S)$; sometimes equivalent to the min-T test
-the average rule: $\varphi_A = \lambda_I + \lambda_S$, or more generally $\varphi_{Ar} = \lambda_I^{w_I} + \lambda_S^{w_S}, w_I, w_S > 0$ (this could be used when the allocation of different weights of importance to sub-hypotheses is required);
-the product-*p* rule: $\varphi_\pi = 1 - (1 - \lambda_I)(1 - \lambda_S)$;
where λ_I and λ_S are the partial p-values related to T_I and T_S respectively. Note that the two p-values are necessarily dependent, thus have to be computed on the same permutation of the data.

In what follows we present more in detail an algorithm for the IU permutation test:

1. read the data set $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = (X_i, i = 1, \ldots, n; n_1, n_2)$ and two limits ε_I and ε_S;

2. define two data vectors $\mathbf{X}_I = (\mathbf{X}_{I1}, \mathbf{X}_{I2}) = (X_{I1i} = X_{1i}, i = 1, \ldots, n_1; X_{I2i} = X_{2i} - \varepsilon_I, i = 1, \ldots, n_2)$ and $\mathbf{X}_S = (\mathbf{X}_{S1}, \mathbf{X}_{S2}) = (X_{S1i} = X_{1i}, i = 1, \ldots, n_1; X_{S2i} = X_{2i} + \varepsilon_S, i = 1, \ldots, n_2)$;

3. compute the observed values of two statistics: $T_I^o = \bar{X}_{I1} - \bar{X}_{I2}$ and $T_S^o = \bar{X}_{S2} - \bar{X}_{S1}$ and take memory;

4. take a random permutation $\mathbf{u}^* = (u_1^*, \ldots, u_n^*)$ of unit labels $\mathbf{u} = (1, \ldots, n)$;

5. define the two permuted data sets: $\mathbf{X}_I^* = (X_{Iu_i^*}, i = 1, \ldots, n; n_1, n_2)$ and $\mathbf{X}_S^* = (X_{Su_i^*}, i = 1, \ldots, n; n_1, n_2)$ both defined on the same permutation $\mathbf{u}^*$;

6. compute the related permuted values of two statistics: $T_I^* = \bar{X}_{I1}^* - \bar{X}_{I2}^*$ and $T_S^* = \bar{X}_{S2}^* - \bar{X}_{S1}^*$ and take memory;

7. independently repeat R times steps 4 to 6 obtaining the results: $[(T^*_{Ir}, T^*_{Sr}),\ r = 1, \ldots, R]$ which simulates the bivariate permutation distribution of two partial tests (T_I, T_S);

8. calculate two estimates of marginal p-value statistics $\lambda_I = \sum_{r=1}^{R} \mathbf{I}[T^*_{Ir} \geq T^o_I]/R$ and $\lambda_S = \sum_{r=1}^{R} \mathbf{I}[T^*_{Sr} \geq T^o_S]/R$ and the φ-combined observed value $\varphi^o = \varphi(\lambda_I, \lambda_S)$, small values of which are evidence against the null hypothesis H;

9. transform the simulated bivariate distribution in step 7 into the bivariate empirical significance level function $\mathbf{L}^* = [(L^*_{Ir}, L^*_{Sr}), r = 1, \ldots, R]$ where $L^*_{hr} = \{0.5 + \sum_{b=1}^{R} \mathbf{I}(T^*_{hb} \geq T^*_{hr})\}/(R+1)$, $h = I, S$;

10. define the φ-combined distribution $[\varphi^*_r = \varphi(L^*_{Ir}, L^*_{Sr}),\ r = 1, \ldots, R]$, that simulates the true bivariate permutation distribution of φ where the dependence between (T_I, T_S) is nonparametrically, albeit implicitly, taken into consideration;

11. the global NPC p-value statistic for testing equivalence is defined as $\lambda_\varphi = \sum_{r=1}^{R} \mathbf{I}[\varphi^*_r \leq \varphi^o]/R$;

12. if $\lambda_\varphi \leq \alpha$ then reject global H in favour of K.

3. The nonparametric UI permutation test

For testing $\widetilde{H}$ against $\widetilde{K}$ within the UI again our proposal is to test separately $\widetilde{H}_I$ against $\widetilde{K}_I$ and $\widetilde{H}_S$ against $\widetilde{K}_S$. To test for $\widetilde{H}_I : \delta_A \geq \delta_B - \varepsilon_I$ against $\widetilde{K}_I : \delta_A < \delta_B - \varepsilon_I \equiv X_{I1} \overset{d}{<} X_{I2}$ we propose the test statistic $\widetilde{T}_I = \bar{X}_{I2} - \bar{X}_{I1}$ and for testing $\widetilde{H}_S : \delta_A \leq \delta_B + \varepsilon_S$ against $\widetilde{K}_S : \delta_A > \delta_B + \varepsilon_S \equiv X_{S1} \overset{d}{>} X_{S2}$ the test statistic $\widetilde{T}_S = \bar{X}_{S2} - \bar{X}_{S1}$ or their permutationally equivalent expressions. Note that even here large values of these statistics are significant, i.e. are evidence of the respective sub-alternatives. The related permutation p-value statistics $\widetilde{\lambda}_I$ and $\widetilde{\lambda}_S$ are now to be calculated in such a way that at least one small p-value statistic is evidence of the global alternative $\widetilde{K}$. It is to be observed here that $\widetilde{K}_I$ true implies $\widetilde{K}_S$ false, and vice versa; whereas $\widetilde{H}_I$ and $\widetilde{H}_S$ can be jointly true. Moreover, now two partial tests $\widetilde{T}_I$ and $\widetilde{T}_S$cannot be jointly unbiased.This fact implies a modification with respect to NPC as in Pesarin and Salmaso (2010) on the combining functions. That is, they must be combined according to a modified NPC methodology, i.e. by combined functions $\psi : [0; 1]^2 \to \mathbb{R}^+$ large values of which are evidence against the global null hypothesis $\widetilde{H}$. Combining function for UI permuation test should satisfy the following properties:

1) ψ is continuous and non-increasing in each argument, i.e. $\widetilde{\lambda}_q < \widetilde{\lambda}'_q$ implies $\psi(\ldots, \widetilde{\lambda}_q, \ldots) \geq \psi(\ldots, \widetilde{\lambda}'_q, \ldots)$;

2) ψ must attain its supremum if at least one argument attains 0;

3) $\alpha > 0$ implies the combined critical value is $\psi_\alpha < \infty$, i.e. no concentration at $+\infty$ under $\widetilde{H}$.

Some possible of combining functions for the UI permutation test are:

-the min-*p* rule: $\psi_m = \max(1 - \widetilde{\lambda}_I, 1 - \widetilde{\lambda}_S)$ corresponding to the so called Tippett's combination rule;

-the product rule: $\psi_P = \widetilde{\lambda}_I \widetilde{\lambda}_S$, equivalent to the famous Fisher's combination rule $\psi_F = -2\left[\ln \widetilde{\lambda}_I + \ln \widetilde{\lambda}_S\right]$;

Let us now introduce in detail an algorithm for the UI permutation test:

1. read the data set $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = (X_i, i = 1, \ldots, n; n_1, n_2)$ and two limits ε_I and ε_S;
2. define two data vectors $\mathbf{X}_I = (\mathbf{X}_{I1}, \mathbf{X}_{I2}) = (X_{I1i} = X_{1i}, i = 1, \ldots, n_1; X_{I2i} = X_{2i} - \varepsilon_I, i = 1, \ldots, n_2)$ and $\mathbf{X}_S = (\mathbf{X}_{S1}, \mathbf{X}_{S2}) = (X_{S1i} = X_{1i}, i = 1, \ldots, n_1; X_{S2i} = X_{2i} + \varepsilon_S, i = 1, \ldots, n_2)$;
3. compute the observed values of two test statistics: $\widetilde{T}_I^o = \bar{X}_{I2} - \bar{X}_{I1}$ and $\widetilde{T}_S^o = \bar{X}_{S1} - \bar{X}_{S2}$ and take memory;
4. take a random permutation $\mathbf{u}^* = (u_1^*, \ldots, u_n^*)$ of unit labels $\mathbf{u} = (1, \ldots, n)$;
5. define the two permuted data sets: $\mathbf{X}_I^* = [X_I(u_i^*), i = 1, \ldots, n; n_1, n_2]$ and $\mathbf{X}_S^* = [X_S(u_i^*), i = 1, \ldots, n; n_1, n_2]$; note that two permuted data sets are both defined on the same permutation $\mathbf{u}^*$;
6. compute the permuted values of two statistics: $\widetilde{T}_I^* = \bar{X}_{I2}^* - \bar{X}_{I1}^*$ and $\widetilde{T}_S^* = \bar{X}_{S1}^* - \bar{X}_{S2}^*$ and take memory;
7. independently repeat R times steps 4 to 6; the results: $[(\widetilde{T}_{Ir}^*, \widetilde{T}_{Sr}^*), r = 1, \ldots, R]$ simulate the bivariate permutation distribution of two partial test statistics $(\widetilde{T}_I, \widetilde{T}_S)$;
8. calculate two estimates of partial p-value statistics $\widetilde{\lambda}_I = \sum_{r=1}^R \mathbf{I}(\widetilde{T}_{Ir}^* \geq \widetilde{T}_I^o)/R$ and $\widetilde{\lambda}_S = \sum_{r=1}^R \mathbf{I}(\widetilde{T}_{Sr}^* \geq \widetilde{T}_S^o)/R$ and the estimated global test statistic $\widetilde{T}_G = \min(\hat{\lambda}_I, \hat{\lambda}_S)$;
9. if $\hat{T}_G > \alpha$ reject the global null hypothesis H_0.

4. A simulation study

In the present section we wish to evaluate the behaviour of the IU and UI permutation tests both under H_0 and in power. We consider three different distributions as data generators, gaussian, uniform and exponential and a two sample design with equal sample size for each sample. The rejection probability of the permutation test (based on $B = 2000$ permutations) at usual nominal significance level of 5% is recorded for different situations on the basis of 2000 Monte Carlo iterations. In particular we consider the situation $\delta_A = \varepsilon_1$ (H_0 both for IU and UI tests), $\delta_A = 0$ (H_0 for UI test, H_1 for IU test) and $\delta_A = 2\varepsilon_2$ (H_0 for IU test, H_1 for UI test). We also consider both the case with symmetrical and asymmetrical equivalence interval. The max-*p* rule and min-*p* rule have been chosen as combining functions respectively for the IU and UI tests. We report in Table 1 and Table 2 the results of the simulations for the case with δ_A on a limit of the equivalence range, i.e. under the null hypothesis both for IU and UI test. The choice of equivalence ranges takes into account the distributions considered. In Table 3 and Table 4 are shown the results for different cases: $\delta_A = 0$ and $\delta_A = 2\varepsilon_2$, and $\delta_A = 0$ and $\delta_A = 2\varepsilon_1$. Note that, we report these results in the same table because if we wish to compare the behaviour of the two tests we have to consider that the case $\delta_A = 0$ is a situation under H_1 for the IU test which corresponds to the case $\delta_A = 2\varepsilon_2$ for the UI test. Similarly for the case with asymmetrical interval we compare the case $\delta_A = 0$ for the IU test with the case $\delta_A = 2\varepsilon_1$ for the UI test.

Table 1: Rejection rates at $\delta_A = \varepsilon_1$(H_0 both for IU and UI tests) with sample size $n_1 = n_2 = 20, 30$ for Gaussian, Uniform and Exponential distribution, and symmetrical equivalence range.

					H_0	
Distribution	n_1	n_2	ε_1	ε_2	IU	UI
Gaussian	20	20	-0.75	0.75	0.049	0.053
Gaussian	30	30	-0.75	0.75	0.045	0.052
Exponential	20	20	-0.50	0.50	0.026	0.048
Exponential	30	30	-0.50	0.50	0.047	0.053
Uniform	20	20	-0.25	0.25	0.051	0.052
Uniform	30	30	-0.25	0.25	0.043	0.045

Table 2: Rejection rates at $\delta_A = \varepsilon_1$(H_0 both for IU and UI tests) with sample size $n_1 = n_2 = 20, 30$ for Gaussian, Uniform and Exponential distribution, and asymmetrical equivalence range.

					H_0	
Distribution	n_1	n_2	ε_1	ε_2	IU	UI
Gaussian	20	20	-0.50	1.00	0.044	0.046
Gaussian	30	30	-0.50	1.00	0.052	0.047
Exponential	20	20	-0.30	0.70	0.036	0.051
Exponential	30	30	-0.30	0.70	0.049	0.042
Uniform	20	20	-0.10	0.40	0.040	0.042
Uniform	30	30	-0.10	0.40	0.054	0.052

Table 3: Rejection rates at $\delta_A = 0$ (H_0 for UI test, H_1 for IU test) and $\delta_A = 2\varepsilon_2$ (H_0 for IU test, H_1 for UI test) with sample size $n_1 = n_2 = 20, 30$ for Gaussian, Uniform and Exponential distribution and symmetrical equivalence range.

					H_1	
					IU	UI
Distribution	n_1	n_2	ε_1	ε_2	$\delta_A = 0$	$\delta_A = 2\varepsilon_2$
Gaussian	20	20	-0.75	0.75	0.507	0.747
Gaussian	30	30	-0.75	0.75	0.770	0.885
Exponential	20	20	-0.50	0.50	0.422	0.510
Exponential	30	30	-0.50	0.50	0.691	0.606
Uniform	20	20	-0.25	0.25	0.712	0.843
Uniform	30	30	-0.25	0.25	0.912	0.949

Table 4: Rejection rates at $\delta_A = 0$ (H_0 for UI test, H_1 for IU test) and $\delta_A = 2\varepsilon_1$ (H_0 for IU test, H_1 for UI test) with sample size $n_1 = n_2 = 20, 30$ for Gaussian, Uniform and Exponential distribution and asymmetrical equivalence range.

					H_1	
					IU	UI
Distribution	n_1	n_2	ε_1	ε_2	$\delta_A = 0$	$\delta_A = 2\varepsilon_1$
Gaussian	20	20	-0.50	1.00	0.388	0.482
Gaussian	30	30	-0.50	1.00	0.582	0.613
Exponential	20	20	-0.30	0.70	0.286	0.261
Exponential	30	30	-0.30	0.70	0.500	0.330
Uniform	20	20	-0.10	0.40	0.284	0.274
Uniform	30	30	-0.10	0.40	0.359	0.385

Note that, under the null hypothesis for the IU test, with an exponential distribution we obtain a conservative behaviour for small sample size ($n_1 = n_2 = 20$) whereas this is not a problem for UI test. Note also that increasing the sample size, things get better both for symmetrical and asymmetrical ranges. It is important to recall that for distribution like the Exponential we have to consider a 'rank-version' of the IU algorithm. This version consists in inserting between step 2 and step 3 of the algorithm presented in section 2, the rank transformation of the X_I and X_S data.

5. Conclusions

Testing for equivalence of two treatments is very common in many area of research such as clinical trials, bioequivalence or pharmaceutical experiments. As pointed out by Sen (Sen, 2007), the critical points in this field are essentially two: the difficulties related to the appropriate application of the likelihood ratio methods and the way to deal with the generally too complex dependence structure of the several partial test statistics where such analysis is usually broken down. In this paper using the nonparametric combination (Pesarin, 2001; Pesarin and Salmaso, 2010) of dependent permutation tests, we have provided two procedures, one following the *intersection-union* (IU) approach and one following the *union-intersection* (UI) approach, able to deal with the intriguing problem of testing for equivalence in a general multidimensional setting. The extension of the proposed procedures to several other designs (i.e. one sample designs, to $C > 2$ samples, ordered categorical endpoint variables, repeated measurements, multidimensional and mixed settings, situations with missing or censored data) is obtained simply by suitably changing the combining functions with respect to the corresponding solutions discussed in Pesarin and Salmaso (2010) and Pesarin and Salmaso (2011). It is important to note that neither the IU nor the UI approaches are uniformly appropriate for all situations, but the choice depends on the specific problem.

References

Berger, R.L. (1982), "Multiparameter hypothesis testing and acceptance sampling", *Technometrics*, 24, 295–300.

Berger, R.L. and Hsu, J.C. (1996), "Bioequivalence trials, intersection-union tests and equivalence confidence sets", *Statistical Science*, 11, 283–319.

Bertoluzzo, F., Pesarin, F. and Salmaso, L. (2013), "On multi-sided permutation tests", *Communications in Statistics - Simulation and Computation*, 42, 1380–1390.

D'Agostino, R.B., Massaro, J.M. and Sullivan, L.M. (2003), "Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics", *Statistics in Medicine*, 22, 169–186.

Hung, H.M.J. and Wang, S.U. (2009), "Some controversial multiple testing problems in regulatory applications", *Journal of Biopharmaceutical Statistics*, 19, 1–11.

Julious, S.A. (2010), *Sample Sizes for Clinical Trials*, Chapman & Hall/CRC, Boca Raton, USA.

Laster, L.L. and Johnson, M.F. (2003), "Non-inferiority trials: the 'at least as good as' criterion", *Statistics in Medicine*, 22, 187–200.

Liu, J-p., Hsueh, H-m., Hsieh, E., and Chen, J.J. (2002), "Tests for equivalence or non-inferiority for paired binary data", *Statistics in Medicine*, 21, 231–245.

Mehta, C.R., Patel, N.R. and Tsiatis, A.A. (1984), "Exact significance testing to establish treatment equivalence with ordered categorical data", *Biometrics*, 40, 819–825.

Pesarin F. (2001), *Multivariate Permutation Tests, with Applications in Biostatistics*, Wiley, Chichester, UK.

Pesarin F. and Salmaso L. (2010), *Permutation Tests for Complex Data, Theory, Applications and Software*, Wiley: Chichester, UK.

Pesarin, F. and Salmaso, L. (2011) "A review and some new results on permutation testing for multivariate problems", *Statistics and Computing*, 22, 639–646.

Roy, S.N. (1953), "On a heuristic method of test construction and its use in multivariate analysis", *The Annals of Mathematical Statistics*, 24, 220–238.

Sen, P.K. (2007), Union–intersection principle and constrained statistical inference, *Statistical Planning and Inference*, 137, 3741–3752.

Wellek, S. (2010), *Testing Statistical Hypotheses of Equivalence and Noninferiority*, Chapman & Hall/CRC: Boca Raton, USA.

Zhong, Z., Chen, W. and Jin, H. (2012), "A new test for testing non inferiority in matched-pairs design", *Communications in Statistics - Simulation and Computation*, 41, 1557-1565.