Statistics
Vol. 00, No. 00, April 2013, 1–19

RESEARCH ARTICLE

Sequential precedence tests

N. Balakrishnan[a*], R. Gouet[b], F.J. López[c] and G. Sanz[c]

[a]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1*; [b]*Departamento de Ingeniería Matemática y Centro de Modelamiento Matemático (UMI 2807, CNRS), Universidad de Chile, Av. Blanco Encalada 2120, 837-0459 Santiago, Chile*; [c]*Departamento de Métodos Estadísticos and BIFI, Facultad de Ciencias, Universidad de Zaragoza. C/ Pedro Cerbuna, 12. 50009 Zaragoza, Spain.*

(*Received 00 Month 200x; in final form 00 Month 200x*)

In this paper we introduce the sequential precedence test for testing the equality of two continuous distribution functions, against the stochastically ordered alternative. This procedure replaces the classical precedence test, used in life-testing experiments, by a sequence of tests which are applied at the failure times of one of the samples. This allows the possibility of stopping the experiment earlier than the precedence test. By means of extensive Monte Carlo simulations and real data, we show that the proposed methodology results in substantial saving of experimental time and cost, without compromising in power. Algorithms for the implementation of the sequential precedence test are included.

Keywords: Precedence statistics; life-testing; sequential monitoring; simulations

AMS Subject Classification: 62G10; 62L10; 62N03

1. Introduction

In life-testing experiments data naturally become available in an ordered fashion, with the smallest observation coming first, the second smallest next, and so on. This feature enables one to apply time-saving statistical procedures, wherein only

*Corresponding author. Email: bala@mcmaster.ca

ISSN: 0233-1888 print/ISSN 1029-4910 online
© 2013 Taylor & Francis
DOI: 10.1080/02331888.YYYY.xxxxxx
http://www.tandfonline.com

early failures need to be observed. In this context, a classical problem is to test for the equality of two continuous distributions, against the stochastically ordered alternative; more specifically,

$$H_0 : F_X = F_Y \quad \text{vs.} \quad H_1 : F_X > F_Y. \tag{1}$$

Precedence tests (PT) are distribution-free, two-sample tests, traditionally used for testing the hypotheses in (1). They are known to be robust, easy to implement and time/cost-efficient. Observe that continuity of the distributions is essential, as otherwise PT would not be distribution-free.

PT are based on the number of observations from one sample that precede a specific order statistic from the other sample. The data from both samples is summarized using the placement statistics $M_1, M_2, \ldots$, where M_i is the number of X-values between the $(i-1)$-th and the i-th Y-order statistics. PT are characterized by a positive integer parameter r, a critical value d and a function $f_r : \mathbb{N}^r \to \mathbb{R}$. The experiment ends when the r-th Y-value is observed; H_0 is rejected if $f_r(M_1, \ldots, M_r) \geq d$. If a given level $\alpha \in (0, 1)$ is specified, then d can be determined so as to make the size of the test equal (or as close as possible) to α.

A well-known example is the classical PT, with $f_r(m_1, \ldots, m_r) = \sum_{i=1}^{r} m_i$, first introduced in [1], based on the exceedance test of [2]. The classical PT has many interesting features but suffers from a tendency to lose power as r increases, for some distributions. This undesirable peculiarity, which can be interpreted as a masking phenomenon, motivated the introduction of the maximal PT in [3] , with $f_r(m_1, \ldots, m_r) = \max\{m_1, \ldots, m_r\}$. Later, the Wilcoxon-type rank-sum PT and other natural extensions of the classical and the maximal PT were introduced; see [4, 5]. More recent developments on the theory and applications of PT can be found in [6], [7] and [8]. The interested reader is referred to the monograph [9], for a thorough discussion of properties of PT and extensive bibliography on the subject.

PT are acknowledged as efficient because they allow to reach a conclusion using relatively few experimental units. In some cases, reducing the number of experimental units is an important ethical or economical issue that has to be addressed.

Our aim is to contribute in this direction by introducing a more cost-efficient sequential-like PT.

A natural measure of efficiency for PT is the user-defined parameter r, which is the number of placement statistics to be observed before coming to a conclusion. In practice, however, a decision could be made earlier if at some stage of the experiment we *anticipate* that H_0 will be rejected. For instance, suppose we have observed $M_1, \ldots, M_i$. Then, if $f_r(M_1, \ldots, M_i, m_{i+1}, \ldots, m_r) \geq d$, for all possible values $m_{i+1}, \ldots, m_r$ of the, so far unobserved, placement statistics $M_{i+1}, \ldots, M_r$, the experiment can be terminated at that point by rejecting H_0. So in fact, a PT can terminate at a random index less than r and the number of experimental units be reduced.

Inspired by the observation above we propose to define a flexible sequential-like PT, resulting in more efficiency and better or comparable power with respect to standard PT. The idea consists in applying a sequence of PTs, as the placement statistics $M_1, M_2, \ldots$ become available, and stopping the experiment at the first rejection of H_0. We call such procedure *sequential precedence test* (SPT). Precise definitions are in Section 2.

It is of interest to mention that our proposal is close in spirit to the sequential monitoring of clinical trials, now a standard practice; see, for example, [10] and [11].

By means of extensive Monte Carlo simulations and real data, we focus on showing that the proposed methodology is advantageous because, in exchange of a slightly more complex implementation, it is more efficient than standard PT. The empirical results support the foregoing claim and allow us to recommend the use of SPT in place of PT.

The rest of the paper is organized as follows. In Section 2 we introduce the notation, give formal definitions and discuss the practical implementation of the tests. Section 3 is dedicated to numerically assess the performance of our methodology, using massive simulation and real data. Results are presented in a series of tables. Concluding remarks are made in Section 4. Finally, the Appendix contains descriptions of the algorithms used in the implementation of the tests, presented in pseudocode style, which could be useful to anyone wishing to analyze their own data.

2. Preliminaries and Definitions

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two mutually independent, iid random samples from continuous distributions F_X and F_Y, respectively, and let $X_{1:n_1} \leq \cdots \leq X_{n_1:n_1}$ and $Y_{1:n_2} \leq \cdots \leq Y_{n_2:n_2}$ be their corresponding order statistics. The placement statistics $M_1, \ldots, M_{n_2}$ are defined as

$$M_i = \#\{j \mid Y_{i-1:n_2} < X_j \leq Y_{i:n_2}\},\ i = 1, \ldots, n_2, \tag{2}$$

where $\#\{\cdot\}$ denotes the cardinality of $\{\cdot\}$ and $Y_{0:n_2} = -\infty$ by convention.

We present below a formal definition of PT. To that end we introduce the notion of test-vector as a vector of functions $\boldsymbol{f} = (f_1, \ldots, f_{n_2})$, whose components $f_i : \mathbb{N}^i \to \mathbb{R}$, $i = 1, \ldots, n_2$, are called test-functions. The set of possible values of $(M_1, \ldots, M_i)$ is denoted by

$$\mathcal{M}_i = \{(m_1, \ldots, m_i) \in \mathbb{N}^i : \sum_{j=1}^{i} m_j \leq n_1\}. \tag{3}$$

Probability and expectation under H_i are denoted by P_i and E_i, respectively, for $i = 0, 1$.

2.1. *Precedence test*

Definition 2.1: Let $r \leq n_2$ be a positive integer, $d \in \mathbb{R}$ and $\boldsymbol{f} = (f_1, \ldots, f_{n_2})$ a test-vector. The PT with placement number r, test-vector $\boldsymbol{f}$ and critical value d (referred to as $(r, \boldsymbol{f}, d)$-PT) has its rejection region as

$$R = \{(m_1, \ldots, m_r) \in \mathcal{M}_r : f_r(m_1, \ldots, m_r) \geq d\}.$$

In practice d is determined so that the size $P_0[R]$ is as close to $\alpha \in (0, 1)$ (from

Table 1. Test-vectors.

Name	i-th component
$\boldsymbol{f}^1$ Classical precedence	$f_i^1 = \sum_{j=1}^{i} m_j$
$\boldsymbol{f}^2$ Maximum	$f_i^2 = \max\{m_1, \ldots, m_i\}$
$\boldsymbol{f}^3$ Wilcoxon-type W_{min}	$f_i^3 = -(n_1(n_1+2i+1)/2 - \sum_{j=1}^{i}(i+1-j)m_j)$
$\boldsymbol{f}^4$ Wilcoxon-type W_{max}	$f_i^4 = -(n_1(n_1+2n_2+1)/2 - \sum_{j=1}^{i}(n_2+1-j)m_j)$
$\boldsymbol{f}^5$ Wilcoxon-type W_E	$f_i^5 = (f_i^3 + f_i^4)/2$
$\boldsymbol{f}^6$ Young's	$f_i^6 = \max_{1\le l\le i}\left\{\sum_{j=1}^{l} m_j - (l-1)\right\}$
$\boldsymbol{f}^7$ Modified Young's	$f_i^7 = \max_{1\le l\le i}\left\{\sum_{j=1}^{l} m_j - \left\lfloor l\frac{n_1}{n_2+1}\right\rfloor\right\}$
$\boldsymbol{f}^8$ Wilcoxon-type W_{mix}	$f_i^8 = f_i^3$ if $i \le \lfloor r/2 \rfloor$; f_i^4 if $i > \lfloor r/2 \rfloor$

below) as possible, that is, the critical value is calculated as

$$d_\alpha = \min\{d : P_0[f(M_1, \ldots, M_r) \ge d] \le \alpha\}. \tag{4}$$

Observe that P_0 is *distribution-free*, in the sense of being independent of the specific form of $F_X(= F_Y)$. So, d_α in (4) can be easily calculated from the null joint probability function of $M_1, \ldots, M_i$, given in page 62 of [9] as

$$P_0[M_1 = m_1, \ldots, M_i = m_i] = \frac{\binom{n_1+n_2-\sum_{j=1}^{i} m_j - i}{n_2 - i}}{\binom{n_1+n_2}{n_2}}, \tag{5}$$

for $(m_1, \ldots, m_i) \in \mathcal{M}_i,\ i = 1, \ldots, r$.

Observe also that R only depends on the r-th component of the test-vector and so it is unnecessary to use a test-vector to characterize the PT. However, we prefer to express it in vector form so that PT is a particular case of SPT, defined below.

We present in Table 1 the test-vectors $\boldsymbol{f}^1, \ldots, \boldsymbol{f}^8$ to be considered in this paper. In the first column we list their names and in the second, their i-th component. We have included functions which are well known and/or have good performance, as reported in the literature; see [9].

The first five test-vectors (functions) in Table 1 have been analyzed in great detail in [9]. Observe that, in contrast to the cited reference, $\boldsymbol{f}^3$ and $\boldsymbol{f}^4$ are expressed here with a leading negative sign in order to have a rejection region of the form $f_r^j(\cdot) \ge d$.

Young's statistic $\boldsymbol{f}^6$ was used by Little in [12] with the purpose of making early decisions in a PT. The statistic originally defined by Young is the instantaneous difference between the number of X-values and the number of Y-values and Little proposed to reject H_0 the first time that Young's statistic attains a critical level d (before the r-th Y-value). Although this test is not formally PT (since it may reject the null hypothesis before r), the rejection region is exactly that of PT, with test-vector $\boldsymbol{f}^6$. Vector $\boldsymbol{f}^7$ is a slight modification of Young's statistic where, instead of subtracting $l-1$ from $\sum_{j=1}^{l} m_j$, we subtract $E_0[\sum_{j=1}^{l} M_j] = ln_1/(n_2+1)$. The test-vector $\boldsymbol{f}^8$ is a sort of mixture between $\boldsymbol{f}^3$ and $\boldsymbol{f}^4$, defined by $f_i^8 = f_i^3$ if $i \leq \lfloor r/2 \rfloor$ and $f_i^8 = f_i^4$ if $i > \lfloor r/2 \rfloor$.

2.1.1. Stopping time of PT

In the naive implementation of PT the experiment always ends when the r-th placement statistic is observed. As commented in the Introduction, this can be significantly improved by anticipating that H_0 will be rejected. After observing each placement statistic, the evidence is evaluated to see if H_0 will be rejected later. This idea is formalized in Definition 2.2.

Let, for $1 \leq i < j \leq r$,

$$\mathcal{M}_{ij} = \Big\{(m_{i+1}, \ldots, m_j) \in \mathbb{N}^{j-i} : \sum_{l=1}^{i} M_l + \sum_{l=i+1}^{j} m_l \leq n_1\Big\}$$

be the (random) set of possible values of $(M_{i+1}, \ldots, M_j)$, given $M_1, \ldots, M_i$, and let also, for $1 \leq i \leq r-1$,

$$\mu_i(M_1, \ldots, M_i) = \min\{f_r(M_1, \ldots, M_i, m_{i+1}, \ldots, m_r) : (m_{i+1}, \ldots, m_r) \in \mathcal{M}_{ir}\},$$

with $\mu_r(M_1, \ldots, M_r) = f_r(M_1, \ldots, M_r)$.

Definition 2.2: The stopping time of the $(r, \boldsymbol{f}, d)$-PT is defined by

$$T = \min\{1 \leq i \leq r : \mu_i(M_1, \ldots, M_i) \geq d\},$$

if $\mu_i \geq d$, for some $i = 1, \ldots, r$, and by $T = r$ otherwise.

2.1.2. *Increasing test-functions*

Definition 2.3: A test-function $f_i : \mathcal{M}_i \to \mathbb{R}$ is said to be increasing if f_i is a non-decreasing function in each of its arguments.

The importance of this property is the simplification it brings to checking the condition $\mu_i(M_1, \dots, M_i) \geq d$. Indeed, for increasing f_r, $\mu_i(M_1, \dots, M_i) = f_r(M_1, \dots, M_i, 0, \dots, 0)$. Although test-functions need not be increasing, a "reasonable" PT has test-vector with increasing functions since it should be easier to reject H_0 when larger values of $M_1, \dots, M_r$ are observed. This is indeed the case of vectors in Table 1.

2.2. *Sequential precedence test*

We introduce the sequential precedence test (SPT) as a more efficient alternative to PT.

Definition 2.4: Let $r \leq n_2$ be a positive integer, $\boldsymbol{d} = (d_1, \dots, d_r) \in \mathbb{R}^r$ and $\boldsymbol{f} = (f_1, \dots, f_{n_2})$ a test-vector. The SPT with placement number r, test-vector $\boldsymbol{f}$ and critical vector $\boldsymbol{d}$ (referred to as $(r, \boldsymbol{f}, \boldsymbol{d})$-SPT) has its rejection region as

$$R_s = \{(m_1, \dots, m_r) \in \mathcal{M}_r : f_i(m_1, \dots, m_i) \geq d_i, \text{ for some } i = 1, \dots, r\}.$$

Given $\alpha \in (0,1)$, the critical vector $\boldsymbol{d}$ should be determined so that the size of the test is closest to $\alpha \in (0,1)$ (from below) but, contrary to standard PT, no simple solution exists in this multidimensional setting. We present below a strategy based on *α-spending sequences*, similar to α-spending functions of clinical trials. Here again, as we deal with placement statistics under H_0, the null joint probability function in (5) can be made use of.

Observe that the $(r, \boldsymbol{f}, d)$-PT is identical to the $(r, \boldsymbol{f}, \boldsymbol{d})$-SPT with $d_r = d$ and $d_1, \dots, d_{r-1}$ large enough so that

$$P_0[f_i(M_1, \dots, M_i) \geq d_i] = 0, \ i = 1, \dots, r-1.$$

This suggests that any PT can be improved, in terms of power and/or cost-efficiency, by a suitably chosen SPT.

2.2.1. Stopping time of SPT

The stopping time of the $(r, \boldsymbol{f}, \boldsymbol{d})$-SPT is the (random) index at which the experiment terminates with a decision. This can be initially interpreted as $\min\{1 \leq i \leq r : f_i(M_1, \ldots, M_i) \geq d_i\}$ if H_0 is rejected, and $T = r$ otherwise. But, as we did for PT, we can examine the available information to see if H_0 will be rejected later on. So, even if $f_j(M_1, \ldots, M_j) < d_j$ for $j = 1, \ldots, i$, H_0 should be rejected at stage i if we foresee that, for some $k > i$, $f_k(M_1, \ldots, M_k) \geq d_k$. More formally, for $1 \leq i < j \leq r$, let

$$\mu_{ij}(M_1, \ldots, M_i) = \min\{f_j(M_1, \ldots, M_i, m_{i+1}, \ldots, m_j) : (m_{i+1}, \ldots, m_j) \in \mathcal{M}_{ij}\}$$

and $\mu_{ii}(M_1, \ldots, M_i) = f_i(M_1, \ldots, M_i)$.

Definition 2.5: The stopping time of the $(r, \boldsymbol{f}, \boldsymbol{d})$-SPT is defined by

$$T_s = \min\{1 \leq i \leq r : \mu_{ij}(M_1, \ldots, M_i) \geq d_j \text{ for some } j \geq i\},$$

if H_0 is rejected, and $T = r$ otherwise.

As in the case of PT, the use of increasing test-functions f_i greatly simplifies the implementation of SPT since $\mu_{ij}(M_1, \ldots, M_i) = f_j(M_1, \ldots, M_i, 0, \ldots, 0)$. Hereinafter we assume that all test functions related to SPT are increasing.

2.2.2. α-spending sequence and critical values

We describe a procedure for calculating a vector of critical values in SPT, using *α-spending sequences*.

Definition 2.6: For $r \leq n_2$ and $\alpha \in (0, 1)$, an α-spending sequence is a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$ such that $\alpha_i \geq 0$, for $i = 1, \ldots, r$, and $\sum_{i=1}^{r} \alpha_i = \alpha$.

Table 2 displays the α-spending sequences $\boldsymbol{\alpha}^j, j = 0, \ldots, 4$, to be considered later in our study. Their names appear in the first column and describe the manner in which the overall α is *spent*. The case $\boldsymbol{\alpha}^0$ corresponds to the PT because α is

Table 2. α-spending sequences.

Name	Notation	Components
Nonsequential	$\boldsymbol{\alpha}^0$	$\alpha_1^0 = \cdots = \alpha_{r-1}^0 = 0,\ \alpha_r^0 = \alpha$
Constant	$\boldsymbol{\alpha}^1$	$\alpha_i^1 = \alpha/r,\ i = 1, \ldots, r$
Linear increasing	$\boldsymbol{\alpha}^2$	$\alpha_i^2 = 2i\alpha/(r(r+1)),\ i = 1, \ldots, r$
Linear decreasing	$\boldsymbol{\alpha}^3$	$\alpha_i^3 = 2(r+1-i)/(r(r+1)), i = 1, \ldots, r$
Extreme bins	$\boldsymbol{\alpha}^4$	$\alpha_1^4 = \alpha/2, \alpha_2^4 = \cdots = \alpha_{r-1}^4 = 0,\ \alpha_r^4 = \alpha/2$

assigned entirely to the last stage.

We now inductively define a vector of critical values as follows.

Definition 2.7: For $r \leq n_2$, let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$ be an α-spending sequence and $\boldsymbol{f}$ a test-vector. Define the critical vector $\boldsymbol{d}_\alpha = (d_{1\alpha}, \ldots, d_{r\alpha})$ inductively as follows. Let

$$d_{1\alpha} = \min\{d : P_0[f_1(M_1) < d] \geq 1 - \alpha_1\}.$$

Assume $d_{1\alpha}, \ldots, d_{(k-1)\alpha}$ given and define $d_{k\alpha}$ as the least d such that

$$P_0\left[\bigcap_{i=1}^{k-1}\{f_i(M_1, \ldots, M_i) < d_{i\alpha}\}, f_k(M_1, \ldots, M_k) < d\right] \geq 1 - \sum_{i=1}^{k} \alpha_i, \qquad (6)$$

for $k = 2, \ldots, r$.

Observe that the $(r, \boldsymbol{f}, \boldsymbol{d}_\alpha)$-SPT related to Definition 2.7 has level α. Observe also that α_i, which can be interpreted as the approximate size of the i-th test differs from the actual probabilities α_i^* of incorrectly rejecting H_0 at stage i but, of course, $\sum_{i=1}^r \alpha_i^* = \sum_{i=1}^r \alpha_i$. Finally notice that, if $\boldsymbol{\alpha} = (0, \ldots, 0, \alpha)$, then the $(r, \boldsymbol{f}, \boldsymbol{d}_\alpha)$-SPT and the $(r, \boldsymbol{f}, d_{r\alpha})$-PT are equivalent.

There seems to be no efficient way of computing $\boldsymbol{d}_\alpha$ from a given α-spending sequence if we require the α_i^*s to stay close to the α_is. This difficulty has no practical implication since the value of α_i must only be understood as a guide. The key point is to keep the size of the test close to but below α. In this respect, the flexibility of SPT allows for the possibility of getting the size closer to the target value α than PT.

The implementation of the $(r, \boldsymbol{f}, \boldsymbol{d}_\alpha)$-SPT requires finding $\boldsymbol{d}_\alpha$ from $\boldsymbol{f}$ and $\boldsymbol{\alpha}$. This can be easily done by using Algorithm 2 (see Appendix), which calls Algorithm 1 for computing the probabilities in (6). Once these values are determined, the implementation of the SPT is described in Algorithm 3.

3. Performance of SPT

We assess the performance of SPT over a wide range of sample sizes, placement numbers, distributions, test-vectors and α-spending sequences. We adopt power and efficiency as the primary criteria in our comparison, with the latter being measured in terms of the expected stopping time of the test; see Definition 3.1 below. Recall that PT and SPT are the same test when $\boldsymbol{\alpha} = (0, \ldots, 0, \alpha)$ and so, comparisons between PT and SPT are included in our study.

Definition 3.1: Consider a SPT with rejection region R_s (Definition 2.4) and stopping time T_s (Definition 2.5). Define

a) the power of the SPT by $P_1[R_s]$,
b) the expected stopping time by $E_1[T_s|R_s]$ and
c) the measure of relative efficiency by $E_1[T_s|R_s]/(r-1)$.

Observe that the expected stopping time is calculated with respect to P_1, conditional on rejection. Observe also that the measure of relative efficiency is normalized by $r-1$ and not by r because the maximal efficiency gain is $r-1$. Finally notice that, unlike P_0, P_1 depends on both F_X and F_Y and so, results concerning power and efficiency, as defined above, are always conditional on the choice of F_X and F_Y.

The SPTs that we consider for evaluation have test-vectors and α-spending sequences as shown in Tables 1 and 2. This gives a total of 40 combinations, 32 of which correspond to genuine SPT (with $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^4$) while 8 are just PT (with $\boldsymbol{\alpha}^0$). We have selected 15 cases to illustrate the computations of the critical vector $\boldsymbol{d}_\alpha$ and the size of the tests. These results, derived exactly from formula (5) (not from simulation), are presented in Table 3. For $\boldsymbol{\alpha}^0$ (corresponding to PT) only $d_{r\alpha}$ is displayed.

For example, in the line corresponding to $\boldsymbol{f}^1$ and $\boldsymbol{\alpha}^4$, under $r = 5$, α is spent

Table 3. Critical vectors $\boldsymbol{d}_\alpha$ and sizes, for $\alpha = .05, n_1 = 15, n_2 = 20$ and $r = 5, 7$.

Test-vector[a]	α-SS	$r = 5$ $\boldsymbol{d}_\alpha = (d_{1\alpha}, \ldots, d_{5\alpha})$	Size	$r = 7$ $\boldsymbol{d}_\alpha = (d_{1\alpha}, \ldots, d_{7\alpha})$	Size
	$\boldsymbol{\alpha}^0$	8	.045	10	.034
	$\boldsymbol{\alpha}^1$	5,6,7,8,9	.044	6,7,7,9,9,10,11	.040
$\boldsymbol{f}^1$	$\boldsymbol{\alpha}^2$	6,7,8,8,8	.047	7,8,8,9,9,10,10	.042
	$\boldsymbol{\alpha}^3$	5,6,7,8,9	.044	5,6,7,8,9,11,11	.049
	$\boldsymbol{\alpha}^4$	5,6,8,10,9	.035	5,6,8,10,11,12,11	.030
	$\boldsymbol{\alpha}^0$	170	.047	184	.046
	$\boldsymbol{\alpha}^1$	130,141,151,161,170	.048	129,140,150,159,168,176,184	.048
$\boldsymbol{f}^3$	$\boldsymbol{\alpha}^2$	129,139,150,160,170	.047	128,138,148,158,167,176,184	.047
	$\boldsymbol{\alpha}^3$	130,141,152,161,169	.046	130,141,151,161,169,176,183	.050
	$\boldsymbol{\alpha}^4$	130,141,150,158,170	.048	130,141,150,158,165,170,183	.045
	$\boldsymbol{\alpha}^0$	6	.030	6	.039
	$\boldsymbol{\alpha}^1$	5,5,5,6,6	.044	6,6,6,6,6,6,6	.039
$\boldsymbol{f}^7$	$\boldsymbol{\alpha}^2$	6,6,6,6,6	.030	7,7,7,7,7,6,6	.039
	$\boldsymbol{\alpha}^3$	5,5,5,6,6	.044	5,5,5,6,6,7,7	.046
	$\boldsymbol{\alpha}^4$	5,5,6,8,6	.040	5,5,6,8,8,8,6	.048

[a]Critical values for $\boldsymbol{f}^3$ (W_{min}) must be taken with a negative sign.

Table 4. Distributions F_X, F_Y: $U[0,1]$ = uniform on $[0,1]$, $Exp(1)$ = standard exponential, $Gamma(1,5)$ = gamma with scale, shape parameters 1,5, $N(0,1)$ = standard normal, $LN(0,0.5)$ = lognormal with parameters $0, 0.5$, $Beta(a,b)$ = beta with parameters a, b.

	F_X	F_Y
Lehmann Type I	$Exp(1)$	$F_X^3(\cdot)$
Shifted Exponential	$Exp(1)$	$F_X(\cdot - 0.5)$
Shifted Gamma	$Gamma(1,5)$	$F_X(\cdot - 0.5)$
Shifted Normal	$N(0,1)$	$F_X(\cdot - 0.5)$
Shifted Lognormal	$LN(0,0.5)$	$F_X(\cdot - 0.5)$
Shifted Uniform	$U[0,1]$	$F_X(\cdot - 0.5)$
Reshaped Beta	$Beta(1,2)$	$1 - F_X(1 - \cdot) = Beta(2,1)$
Lehmann Type II	$Exp(1)$	$1 - (1 - F_X(\cdot))^2$

half in the first stage and half in the last. The values $5, 6, 8, 10, 9$ mean that H_0 is rejected in the first stage if $M_1 \geq 5$; in the second if $M_1 < 5$ and $M_1 + M_2 \geq 6$, etc. Observe that, in spite of the value $d_{4\alpha} = 10$, H_0 is to be rejected at stage 4 if $M_1 + \cdots + M_4 \geq 9$ because of the *anticipative* implementation of the test. Finally, note that the size is 0.035, less than the target value 0.05. The lines corresponding to $\boldsymbol{\alpha}^0$ are standard PT and values there coincide with those reported in [9].

Our objective is to evaluate SPTs over an ample selection of distributions, so that conclusions are useful for a range of applications. We have chosen to work with pairs F_X, F_Y shown in Table 4. The first five have been studied in [9], in the context of PT, and we have added three more to extend the scope of the study. Observe that in 5 out of 8 cases, the distribution F_Y is a shifted version of F_X.

Table 5. Power and expected stopping time, for $n_1 = 15$, $n_2 = 20$, $r = 5$ and $\alpha = 0.05$.

		$\boldsymbol{f}^1$		$\boldsymbol{f}^3$		$\boldsymbol{f}^7$	
F_X, F_Y	α-SS	Power	EST	Power	EST	Power	EST
	$\boldsymbol{\alpha}^0$	0.770	2.83	0.875	1.68	0.756	1.99
Lehmann Type I	$\boldsymbol{\alpha}^1$	0.832	1.60	0.877	1.63	0.832	1.60
	$\boldsymbol{\alpha}^2$	0.800	2.15	0.875	1.68	0.756	1.99
	$\boldsymbol{\alpha}^0$	0.606	2.37	0.891	1.17	0.728	1.21
Shifted Exponential	$\boldsymbol{\alpha}^1$	0.858	1.07	0.897	1.14	0.858	1.07
	$\boldsymbol{\alpha}^2$	0.749	1.31	0.891	1.17	0.728	1.21
	$\boldsymbol{\alpha}^0$	0.363	3.70	0.465	2.32	0.321	2.74
Shifted Gamma	$\boldsymbol{\alpha}^1$	0.407	2.15	0.469	2.22	0.407	2.15
	$\boldsymbol{\alpha}^2$	0.385	3.11	0.465	2.32	0.321	2.74
	$\boldsymbol{\alpha}^0$	0.292	3.94	0.306	2.63	0.240	3.20
Shifted Normal	$\boldsymbol{\alpha}^1$	0.293	2.62	0.308	2.54	0.293	2.62
	$\boldsymbol{\alpha}^2$	0.302	3.57	0.306	2.63	0.240	3.20
	$\boldsymbol{\alpha}^0$	0.487	3.37	0.653	1.99	0.467	2.30
Shifted Lognornal	$\boldsymbol{\alpha}^1$	0.578	1.79	0.659	1.89	0.578	1.79
	$\boldsymbol{\alpha}^2$	0.526	2.61	0.653	1.99	0.467	2.30
	$\boldsymbol{\alpha}^0$	0.945	1.69	0.992	1.05	0.965	1.13
Shifted Uniform	$\boldsymbol{\alpha}^1$	0.986	1.04	0.992	1.04	0.986	1.04
	$\boldsymbol{\alpha}^2$	0.973	1.16	0.992	1.05	0.965	1.13
	$\boldsymbol{\alpha}^0$	0.396	3.83	0.387	2.61	0.335	3.17
Reshaped Beta	$\boldsymbol{\alpha}^1$	0.389	2.65	0.389	2.54	0.389	2.65
	$\boldsymbol{\alpha}^2$	0.406	3.49	0.387	2.61	0.335	3.17
	$\boldsymbol{\alpha}^0$	0.314	4.00	0.280	2.79	0.255	3.40
Lehmann Type II	$\boldsymbol{\alpha}^1$	0.294	2.90	0.281	2.72	0.294	2.90
	$\boldsymbol{\alpha}^2$	0.320	3.73	0.280	2.79	0.255	3.40

3.1. *Results*

While formula (5) is available to calculate $P_0[R_s]$, there is no closed-form expression for the power $P_1[R_s]$, except in the case of Lehmann Type I and Type II distributions. Further, no closed-form exists for the expected stopping time. Therefore, we estimate these values by means of Monte Carlo simulations, where, for each combination, a total of 10^5 runs were carried out.

3.1.1. Preliminary evaluation

We consider three test-vectors: $\boldsymbol{f}^1$ (classical precedence), $\boldsymbol{f}^3$ (W_{min}) and $\boldsymbol{f}^7$ (modified Young's). For the α-spending sequences, besides the baseline $\boldsymbol{\alpha}^0$ of PT, we take $\boldsymbol{\alpha}^1$ (constant) and $\boldsymbol{\alpha}^2$ (linear increasing); see Tables 1 and 2 for details. The sample sizes are set as $n_1 = 15$ and $n_2 = 20$, and the placement number as $r = 5$; other choices of n_1, n_2, r will be considered later on. Combining the above settings with all pairs of distributions from Table 4, we have 72 cases in total. Table 5 presents the results, showing the estimates of power (under the heading Power) and of the expected stopping time (under the heading EST).

As can be seen from Table 5, SPT outperforms PT in terms of both power and early rejection. This is clearly seen for $\boldsymbol{f}^1$ while, for $\boldsymbol{f}^3$ and $\boldsymbol{f}^7$, the performance of SPT is slightly better than that of PT, when using $\boldsymbol{\alpha}^1$. On the other hand, PT and SPT perform as well when $\boldsymbol{\alpha}^2$ is used.

3.1.2. Extended evaluation

We expand the scope by considering multiple choices, which define simulation scenarios. A scenario is defined in terms of n_1, n_2, r and the distribution pair F_X, F_Y. Observe that r has to be chosen in accordance with n_1 and n_2 and, in this respect, we follow [9]. Our choices, detailed below, yield a total of 96 scenarios.

- $(n_1, n_2) \in \{(10, 10), (10, 15), (15, 15), (15, 20)\}$;
- $r \in \{3, 4, 5\}$ if $(n_1, n_2) \in \{(10, 10), (10, 15), (15, 15)\}$,
- $r \in \{3, 5, 7\}$ if $(n_1, n_2) = (15, 20)$;
- All pairs F_X, F_Y from Table 4.

Within each scenario we use every test-vector $\boldsymbol{f}^j, j = 1, \ldots, 8$, from Table 1, and compare PT (related to $\boldsymbol{\alpha}^0$) with SPT (related to $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^4$). Thus, for each $\boldsymbol{f}^j$ we make $384 = 96 \times 4$ comparisons.

3.1.3. Percentages

Results in terms of percentages are presented in Table 6, where each row is associated with a test function, from $\boldsymbol{f}^1$ to $\boldsymbol{f}^8$. Column 1 displays the percentage of the 384 comparisons where PT and SPT are identical; column 2 the percentage of cases where the size of SPT is larger (closer to α) than that of PT; column 3 the percentage of cases where SPT has greater power than PT and column 4 the percentage of cases where SPT has smaller expected stopping time (under H_1) than PT. Note that in columns 2, 3 and 4 percentages are calculated with respect to the number of cases where PT and SPT differ. The computational implementation leading to Table 6 is based on Algorithms 1, 2 and 3, shown in the Appendix. Results in the first two columns are exact while those in the last two are estimated by Monte Carlo simulations, with 10^5 runs for each case.

For the sake of illustration, in row 1, column 1, we read that PT and SPT are the same test, in 2.08% of the instances, while in row 2, column 1 we see they coincide 70.83% of times. In column 4, rows 1, 3, 6 and 7, we find that, for test-vectors $\boldsymbol{f}^1, \boldsymbol{f}^3, \boldsymbol{f}^6$ and $\boldsymbol{f}^7$, SPT rejects H_0 earlier (on average) than PT, in 100% of cases.

Table 6. Percentages.

	Equal tests	Greater size	Higher power	Earlier stop
$\boldsymbol{f}^1$	2.08	87.23	94.15	100
$\boldsymbol{f}^2$	70.83	92.86	97.32	96.43
$\boldsymbol{f}^3$	54.17	77.27	81.82	100
$\boldsymbol{f}^4$	2.08	76.60	82.18	98.94
$\boldsymbol{f}^5$	8.33	84.09	80.11	97.44
$\boldsymbol{f}^6$	64.58	94.12	95.59	100
$\boldsymbol{f}^7$	47.92	96.00	96.50	100
$\boldsymbol{f}^8$	0	62.50	79.17	99.22

Table 7. Power increase.

	Min	Q_1	Med	Q_3	Max
$\boldsymbol{f}^1$	-0.065	0.039	0.073	0.136	0.479
$\boldsymbol{f}^2$	-0.030	0.057	0.099	0.156	0.241
$\boldsymbol{f}^3$	-0.029	0.001	0.006	0.015	0.098
$\boldsymbol{f}^4$	-0.041	0.003	0.019	0.042	0.259
$\boldsymbol{f}^5$	-0.037	0.002	0.013	0.029	0.193
$\boldsymbol{f}^6$	-0.068	0.033	0.066	0.107	0.276
$\boldsymbol{f}^7$	-0.068	0.033	0.057	0.096	0.276
$\boldsymbol{f}^8$	-0.050	0.002	0.026	0.056	0.282

3.1.4. Beyond percentages

We look at differences of power and efficiency by computing the increase (possibly negative) of power and of relative efficiency of SPT with respect to PT. For each test-vector $\boldsymbol{f}^j, j = 1, \ldots, 8$, the power difference is calculated as the power of SPT using $\boldsymbol{\alpha}^j, j = 1, \ldots, 4$, minus the power using $\boldsymbol{\alpha}^0$ (actually, PT). In these calculations we take into account only combinations (out of the 384 considered) where PT and SPT are not equal. Analogously, the difference of relative efficiency is obtained by subtracting the value related to $\boldsymbol{\alpha}^0$ from the one related to $\boldsymbol{\alpha}^j, j = 1, \ldots, 4$.

In Tables 7 and 8, we display the minimum Min, the first quartile Q_1, the median Med, the third quartile Q_3 and the maximum Max of the quantities described above. For instance, the value 0.259 in row 4, column 5 of Table 7 means that, among all combinations there is one where the difference between SPT and PT in terms of power is 0.259, when $\boldsymbol{f}^4$ is used. Note that we summarize results using order statistics, instead of averages and standard deviations, because of the heterogeneity of the instances.

In Tables 6-8 we find that in most combinations where SPT and PT differ, SPT has greater power and efficiency, which can be high in some cases. Even when SPT has lower power, the difference is quite small and clearly compensated by the

Table 8. Increase of relative efficiency.

	Min	Q_1	Med	Q_3	Max
$\boldsymbol{f}^1$	3.21	22.97	31.29	39.72	69.09
$\boldsymbol{f}^2$	-0.57	0.41	3.68	8.92	17.66
$\boldsymbol{f}^3$	0.21	5.22	8.79	12.01	21.67
$\boldsymbol{f}^4$	-0.41	11.74	18.36	26.24	42.46
$\boldsymbol{f}^5$	-1.43	7.04	15.18	21.39	36.76
f^6	1.13	7.39	9.89	12.50	18.95
$\boldsymbol{f}^7$	2.00	7.90	10.80	14.08	22.56
$\boldsymbol{f}^8$	-0.37	14.05	20.23	26.89	42.08

efficiency gain.

A few negative entries appear in the Min column of Table 8, which means that, for some combinations, the expected stopping time increases. This is counterintuitive since SPT is designed to stop earlier than the PT. We believe this happens mainly because SPT may reject H_0 in the last stages, while PT does not reject H_0, thus shifting the stopping time of SPT to the right. Nevertheless, this phenomenon is indeed rarely seen: out of 384 combinations (for each test vector) it only happens in 4 of them, for $\boldsymbol{f}^2$ and $\boldsymbol{f}^4$; in 8, for $\boldsymbol{f}^5$, and in 3, for $\boldsymbol{f}^8$. It does not occur in any combination for $\boldsymbol{f}^1$, $\boldsymbol{f}^3$, $\boldsymbol{f}^6$ and $\boldsymbol{f}^7$. Moreover, the negative values in Table 8 are quite small (recall they are percentages) and this indicates that the unnormalized values (differences of expected stopping times) are very close to 0, the maximum being 0.043.

It is clear from Tables 6 and 8 that, in most cases, a decrease of the expected stopping time is observed. Furthermore, an important reduction can be achieved in the total number of placement statistics needed until rejection: medians range from 3.68%, for $\boldsymbol{f}^2$, to 31.29%, for $\boldsymbol{f}^1$.

3.1.5. Optimal α-spending sequence

An important issue in our methodology is the proper choice of the α-spending sequence. It is worth mentioning here that the analogous problem is also a matter of discussion in the field of clinical trials but, to our knowledge, no clear answer has emerged so far.

A guide for an optimal α-spending sequence is out of the scope of this paper. We can expect such optimum to depend quite strongly on the shapes of F_X or F_Y and so, a universal and useful solution seems infeasible. Instead, we explore a simple strategy, independent of F_X or F_Y, whereby we define as optimal the α-spending

Table 9. Percentages under optimal α-spending sequence.

	Equal tests	Greater size	Higher power	Earlier stop
$\boldsymbol{f}^1$	0	100	98.96	100
$\boldsymbol{f}^2$	41.67	100	100	92.86
$\boldsymbol{f}^3$	41.67	100	94.64	100
$\boldsymbol{f}^4$	0	83.33	83.33	98.96
$\boldsymbol{f}^5$	0	100	81.25	98.96
$\boldsymbol{f}^6$	41.67	100	100	100
$\boldsymbol{f}^7$	16.67	100	100	100
$\boldsymbol{f}^8$	0	83.33	80.21	98.96

Table 10. Power increase under optimal α-spending sequence.

	Min	Q_1	Med	Q_3	Max
$\boldsymbol{f}^1$	-0.012	0.054	0.094	0.161	0.458
$\boldsymbol{f}^2$	0.041	0.061	0.112	0.161	0.241
$\boldsymbol{f}^3$	-0.009	0.006	0.014	0.027	0.095
$\boldsymbol{f}^4$	-0.041	0.004	0.027	0.051	0.231
$\boldsymbol{f}^5$	-0.037	0.004	0.015	0.038	0.187
$\boldsymbol{f}^6$	0.022	0.040	0.081	0.119	0.221
$\boldsymbol{f}^7$	0.014	0.040	0.068	0.105	0.221
$\boldsymbol{f}^8$	-0.050	0.003	0.031	0.070	0.282

Table 11. Increase of relative efficiency using optimal α-spending sequence.

	Min	Q_1	Med	Q_3	Max
$\boldsymbol{f}^1$	3.21	25.84	34.46	41.47	68.93
$\boldsymbol{f}^2$	-0.57	0.30	3.17	8.24	11.35
$\boldsymbol{f}^3$	0.21	7.10	11.14	13.96	20.36
$\boldsymbol{f}^4$	-0.37	16.33	24.29	30.07	41.03
$\boldsymbol{f}^5$	-0.37	12.60	20.90	24.82	36.76
$\boldsymbol{f}^6$	1.13	7.22	9.62	11.03	14.70
$\boldsymbol{f}^7$	2.00	7.80	10.13	13.01	22.56
$\boldsymbol{f}^8$	-0.37	15.36	23.00	29.44	40.84

sequence leading to a test with largest size.

We have implemented this procedure on the combinations considered above. Recall that we have 384 for each test-vector, which correspond to 96 possibilities of (n_1, n_2, r) and F_X, F_Y, for every α-spending sequence $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^4$. In each of the 96 combinations, we choose $\boldsymbol{\alpha}^*$ (among $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^4$) such that the size of the corresponding test is closest to $\alpha = 0.05$. Table 9 (with the same structure of Table 6) presents percentages of better performance of SPT, when $\boldsymbol{\alpha}^*$ is used. Tables 10 and 11 show results analogous to those in Tables 7 and 8, respectively.

Results in Tables 9-11 show an improvement, especially in terms of power, over those in Tables 6-8, indicating that the suggested choice of spending sequence is quite reasonable. Observe in the third column of Table 9 that power increases (with respect to PT) from 80% to 100% of cases, depending on the test-vector. Also, from Table 10, we learn that the median of power increase ranges from 0.014 to 0.112.

Table 12. Performace on Nelson's data, with $n_1 = n_2 = 10, r = 6$ and $\alpha = 0.05$.

	$\boldsymbol{\alpha}^0$	$\boldsymbol{\alpha}^1$	$\boldsymbol{\alpha}^2$	$\boldsymbol{\alpha}^3$	$\boldsymbol{\alpha}^4$	$\boldsymbol{\alpha}^*$
$\boldsymbol{f}^1$	(–, 0.016)	–	1	–	–	(–, 0.050)
$\boldsymbol{f}^2$	(–, 0.032)	–	1	–	–	(1, 0.043)
$\boldsymbol{f}^3$	(4, 0.045)	4	1	4	1	(1, 0.050)
$\boldsymbol{f}^4$	(6, 0.049)	6	1	6	6	(6, 0.050)
$\boldsymbol{f}^5$	(6, 0.049)	6	1	6	6	(6, 0.050)
$\boldsymbol{f}^6$	(–, 0.026)	–	1	–	–	(1, 0.035)
$\boldsymbol{f}^7$	(–, 0.026)	–	1	–	–	(1, 0.035)
$\boldsymbol{f}^8$	(6, 0.049)	4	1	6	1	(1, 0.049)

3.1.6. Real data

We close with a performance study of SPT on real data. The data were taken from Nelson's book [13] and correspond to times to insulating fluid breakdown. See pages 58, 99 and 172 in [9] where these data were used to analyze the performance of PT. The sample sizes are $n_1 = n_2 = 10$; for the placement number, as in [9], we take $r = 6$ and the level is set to $\alpha = 0.05$. The observed values of the placement statistics are

$$(M_1, \ldots, M_6) = (5, 0, 0, 2, 0, 2).$$

Results are displayed in Table 12, where rows correspond to test-vectors $\boldsymbol{f}^j, j = 1, \ldots, 8$ and columns to α-spending sequences $\boldsymbol{\alpha}^j, j = 0, \ldots, 4$ and $\boldsymbol{\alpha}^*$, defined in the previous paragraph. The first columns shows the performance of PT, in terms of the stopping time and size (first and second values between parentheses). The last column, associated to the optimal α-spending sequence, has the same description as the first one. All other columns, for $\boldsymbol{\alpha}^j, j = 1, \ldots, 4$, show just the value of the stopping time. Symbol – indicates that H_0 was not rejected. For example, from line 6 we learn that (nonsequential) Little's test has a size of 0.026 and does not reject H_0. But, when $\boldsymbol{\alpha}^2$ or $\boldsymbol{\alpha}^*$ are used, H_0 is rejected (with stopping time 1) and, for the latter, the size attains the value 0.035.

In the example with Nelson's data we see that the size of SPT is closer to α than PT. This leads to rejection of H_0 by SPT in cases where PT, with size far from α, does not reject. Also, a significant saving in experimental units is achieved.

4. Concluding remarks

In light of the results obtained from simulation and real data, we see that SPT is more efficient than PT in terms of saving experimental units. This saving is not translated into a loss of power; on the contrary, since SPT has size closer to α than PT, its power increases in most cases or shows a very small decrease, in a few occasions. Moreover, SPT is easy to implement. We therefore conclude that SPT is superior and should be preferred to PT in life-testing experiments.

It is worth noticing that PT have also been developed for testing hazard rate ordering between two distributions; see, for example, [14, 15]. In this regard, it would be of interest to explore the possibility of developing sequential-type versions for these procedures.

Acknowledgements

We thank two anonymous referees for their careful reading of the manuscript and for suggestions that helped us to improve the paper. This research was supported by PFB-03-CMM project, Fondecyt grant 1120408 and project MTM2010-15972 of MICINN and FEDER. The last three authors are members of the research group Modelos Estocásticos (DGA).

References

[1] L.S. Nelson, *Tables of precedence life tests*, Technometrics 5 (1963), pp. 491–499.

[2] B. Epstein, *Comparison of some nonparametric tests against normal alternatives with an application to life testing*, J. Am. Stat. Assoc. 50 (1955), pp. 894–900.

[3] N. Balakrishnan and R. Frattina, *Precedence test and maximal precedence test*, in *Recent Advances in Reliability Theory: Methodology, Practice and Inference*, N. Limnios and M. Nikulin, eds., Birkhäuser, Boston, 2000, pp. 355–378.

[4] H.K.T. Ng and N. Balakrishnan, *Wilcoxon-type rank-sum precedence tests*, Aust. N.Z. J. Stat. 46 (2004), pp. 631–648.

[5] ———, *Weighted precedence and maximal precedence tests and an extension to progressive censoring*, J. Stat. Plan. Infer. 135 (2005), pp. 197–221.

[6] N. Balakrishnan, A. Dembińska, and A. Stepanov, *Precedence-type tests based on record values*, Metrika 68 (2008), pp. 233–255.

[7] N. Balakrishnan, R.C. Tripathi, N. Kannan, and H.K.T. Ng, *Some nonparametric precedence-type*

tests based on progressively censored samples and evaluation of power, J. Stat. Plan. Infer. 140 (2010), pp. 559–573.

[8] A. Jordan and B.G. Ivanoff, *Multidimensional plots and precedence tests for point processes on* $\mathbb{R}^d$, J. Multivariate Anal. 115 (2013), pp. 122–137.

[9] N. Balakrishnan and H.K.T. Ng *Precedence-Type Tests and Applications*, Wiley, New York, 2006.

[10] C. Jennison and B.W. Turnbull *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall, Boca Raton, FL., 2000.

[11] H. Zhu and F. Hu, *Sequential monitoring of response-adaptive randomized clinical trials*, Ann. Stat. 28 (2010), pp. 2218–2241.

[12] R. Little, *Tables for making an early decision in precedence tests*, Journal of Testing and Evaluation 2 (1974), pp. 84–86.

[13] W. Nelson *Applied Life Data Analysis*, Wiley, New York, 1982.

[14] M. Sharafi, N. Balakrishnan, and B.E. Khaledi, *Distribution-free comparison of hazard rates of two distributions under Type-II censoring*, Comm. Statist. – Theor. Meth. 42 (2013), pp. 1889–1898.

[15] ———, *Distribution-free comparison of hazard rates of two distributions under progressive type-II censoring*, J. Stat. Comput. Simul. 83 (2013), pp. 1527–1542.

Appendix A. Algorithms

Algorithm *A.1* Function *prob_region_i*

Require: $n_1, n_2, f_1, \ldots, f_i, d_1, \ldots, d_i$
Ensure: $P[f_1(M_1) < d_1, f_2(M_1, M_2) < d_2, \ldots, f_i(M_1, \ldots, M_i) < d_i \mid H_0]$

```
A ← 0
for all j_1 = 0, ..., n_1 do
  if f_1(j_1) < d_1 then
    for all j_2 = 0, ..., n_1 − j_1 do
      if f_2(j_1, j_2) < d_2 then
        ...
        for all j_i = 0, ..., n_1 − j_1 − ··· − j_{i−1} do
          if f_i(j_1, ..., j_i) < d_i then
            s ← \sum_{l=1}^{i} j_l
            A ← A + \binom{n_1+n_2−s−i}{n_2−i} / \binom{n_1+n_2}{n_2}
          end if
        end for
        ...
      end if
    end for
  end if
end for
RETURN A
```

Algorithm *A.2* Computation of critical values $d_1, \ldots, d_r$

Require: $n_1, n_2, f_1, \ldots, f_r, \alpha_1, \ldots, \alpha_r, x_1, \ldots, x_r$
Ensure: $d_1, \ldots, d_r$

```
α'_1 ← α_1
for all i = 2, ..., r do
  α'_i ← α'_{i−1} + α_i
end for
for all i = 1, ..., r do
  d_i ← x_i + 1
  while prob_region_i(d_1, ..., d_i) ≥ 1 − α'_i do
    d_i ← d_i − 1
  end while
  d_i ← d_i + 1
end for
RETURN d_1, ..., d_r
```

Algorithm *A.3* Test

Require: $d_1, \ldots, d_r, m_1, \ldots, m_r$
Ensure: Decision, stopping time

```
for all i = 1, ..., r do
  for all j = i, ..., r do
    if f_j(m_1, ..., m_i, 0, ..., 0) ≥ d_j then
      ending-bin= i
      RETURN "Reject H_0", stopping time
    end if
  end for
end for
ending-bin=r
RETURN "Do not reject H_0", stopping time
```