Goodness-of-Fit Tests Based on Sample Space Partitions: a Unifying Overview

O. THAS[†] AND J. P. OTTOY   olivier.thas@rug.ac.be
Department of Applied Mathematics Biometrics and Process Control Ghent University, B-9000 Gent, Belgium

Abstract. Recently the authors have proposed tests for the one-sample and the k-sample problem, and a test for independence. All three tests are based on sample space partitions, but they were originally developed in different papers. Here we give an overview of the construction of these tests, stressing the common underlying concept of "sample space partitions."

Keywords: goodness-of-fit, rank statistics, sample space partition.

1.   Introduction

In a series of papers the authors (Thas and Ottoy [18], [19], [20]) presented new tests for testing one-sample goodness-of-fit, k-sample goodness-of-fit and independence between two continuous variables. All tests are based on sample space partitions (SSP) of an arbitrary size. They are referred to as the SSPc, SSPkc and SSPrc tests respectively. The aim of this paper is to give an overview of the construction of the test statistics. The stress is laid on the common underlying idea of observation-based SSP's, which categorize the data into classes on which Pearson χ^2-statistics are calculated.

Nowadays it is generally known that the k-sample problem and the problem of independence are both special cases of the traditional goodness-of-fit problem. One of the oldest solutions to the one-sample goodness-of-fit problem is to group, i.e. to categorize, the continuous data into classes (or cells), and next to apply a Pearson χ^2 goodness-of-fit test to the induced discrete data set. Major questions to this approach include where to place the cell boundaries and how many cells must be constructed. Starting with Fisher [4], up to the late 1980's many papers have been devoted to these questions. A summary is given by e.g. Moore [7]. Despite the intuitive justification of this approach, it is recommended these days to apply only

---

[†] Requests for reprints should be sent to O. Thas, Department of Applied Mathematics Biometrics and Process Control Ghent University, B-9000 Gent, Belgium.

goodness-of-fit methods that are specifically developed for continuous data, e.g. a Kolmogorov-Smirnov or a Shapiro-Wilk test for testing normality. Still there are good methods that rely heavily on the Pearson statistic for discrete data. For example, for the one-sample and the k-sample problem, the Anderson-Darling type test statistics (Anderson and Darling [1]; Pettit [8]; Scholtz and Stephens [11]) consist of averages of one-degree-of-freedom Pearson statistics, calculated on all one or two way tables that are induced by observation-centred sample space space partitions. For the k-sample and the independence problem, the other extreme end solution has recently been proposed by Rayner and Best [10]. Their approach consists of identifying a cell boundary with each observation in the sample, resulting in a maximal degrees-of-freedom Pearson statistic of which they consider the first few components as interesting test statistics.

Our approach is situated somewhere in the middle between the Anderson-Darling and the Rayner and Best methods. In particular we will repeatedly group the data into an arbitrary number of classes (fixed *a priori*) according to observation-based cell boundaries. Later we will show that this classification is related to sample space partitions, and we will show that the number of groups is closely related to the SSP-size. The number of different possible groupings depends on the SSP-size. On each differently categorized sample a Pearson statistic is calculated and the SSP test statistics are basically the averages of all such possible Pearson statistics. In the next section, the construction is explained in detail. The small sample power characteristics are discussed in Section 3. Finally, in Section 4 an example of the one-sample SSPc test is presented.

2.   The Sample Space Partition Tests

The Three Goodness-of-Fit Problems

In general all tests may be considered in some sense as goodness-of-fit tests, i.e. they test a null hypothesis that may be written as

$$H_0 : F(x) = G(x) \text{ for all } x \in \mathcal{S}, \tag{1}$$

where $\mathcal{S}$ is the sample space of X and F and G are the true and the hypothesized distribution function, respectively. Depending on how G is determined, different specific types of goodness-of-fit problems arise. In particular, the null hypothesis in (1) is the null hypothesis of the one-sample goodness-of-fit problem. A distinction must be made between the case where G is completely known, and the case where G is known up to some unknown nuisance parameters $\boldsymbol{\theta}$. These cases are typically referred

to as a simple and a composite null hypothesis, respectively. In the latter situation the hypothesized distribution is often denoted by G_θ, where the unknown finite dimensional parameter $\boldsymbol{\theta}$ is estimated from the data, say by means of a consistent estimator $\hat{\boldsymbol{\theta}}_n$, resulting in $\hat{G} = G_{\hat{\theta}}$. For the sake of generality, we will denote the hypothesized distribution G always as $\hat{G}$, even when the null hypothesis is simple.

For the k-sample problem, the null hypothesis is

$$H_0 : F_1(x) = \ldots = F_k(x) \text{ for all } x \in \mathcal{S}, \qquad (2)$$

where the F_i are the true distribution functions of k populations. Under this null hypothesis the common distribution function is denoted by $G(x) = F_1(x) = \ldots = F_k(x)$. The common distribution G is not known and is typically estimated by the empirical distribution function, $\hat{G}$, by pooling all k samples.

Suppose that $\boldsymbol{X} = (Y, Z)$ denotes a bivariate variable. The null hypothesis of independence between Y and Z implies the restriction on the distribution function F_{yz} that it is the product of the univariate marginal distribution functions F_y and F_z of Y and Z, respectively, i.e. $G(y, z) = F_y(y)F_z(z)$ for all (y, z) in the sample space of $\boldsymbol{X}$. The univariate marginal distributions are not specified *a priori*, and thus they must be estimated from the data. The corresponding empirical distribution functions are denoted by $\hat{F}_y$ and $\hat{F}_z$, resulting in the estimated bivariate distribution $\hat{G}(y, z) = \hat{F}_y(y)\hat{F}_z(z)$ under the null hypothesis.

The Observation-Based Sample Space Partitions

Although all three types of SSP tests are based on partitions of the sample space $\mathcal{S}$, we still have to make a distinction between these three types w.r.t. the way the partitions are constructed. First some notation is introduced. Let $\mathcal{S}_n$ denote a sample of n observations. In the case of the k-sample problem and the independence problem, the sample is considered as the pooled sample and the sample of bivariate observations, respectively. For the former problem, the k independent samples are denoted by $\mathcal{S}_{j,n_j}$ ($j = 1, \ldots, k$), where n_j is the corresponding sample size, and $n = \sum_{j=1}^{k} n_j$. For the independence problem we introduce $\mathcal{S}_n^y$ and $\mathcal{S}_n^z$ as the sample of the univariate Y and Z components of the sample $\mathcal{S}_n$. Finally, $\hat{F}_n$ and $\hat{F}_{j,n_j}$ denote the empirical distribution functions as estimators of F and F_j, respectively.

- *One-sample problem.* Apart from the specification of $\hat{G}$, there is at this point no need to treat the simple and the composite null hypothesis case separately. Suppose that D_c is a subset of the sample $\mathcal{S}_n$ with

$c - 1$ elements. Let $\mathcal{D}_c$ denote the set of all such subsets, and let $d_{c,n} = \#\mathcal{D}_c = \binom{n}{c-1}$. Let $x_{(i)} \in D_c$ denote the i-th order statistic of D_c $(i = 1, \ldots, c-1)$, and define $x_{(0)} = \inf \mathcal{S}$ and $x_{(c)} = \sup \mathcal{S}$. We then define a SSP of size c as $[A](D_c) = \{A_1(D_c), \ldots, A_c(D_c)\}$, where $A_i(D_c) = \{x \in \mathcal{S} : x_{(i-1)} < x \leq x_{(i)}\}$ $(i = 1, \ldots, c)$. The partition $[A](D_c)$ implies on the sample $\mathcal{S}_n$ a discretization into a table with c cells of which the counts are given by $N_i(D_c) = \#(A_i(D_c) \cap \mathcal{S}_n) = n \int_{A_i} d\hat{F}_n$, $i = 1, \ldots, c$. The construction of the partitions and the induced table of counts is illustrated in Figure 1.

- *The k-sample problem.* Suppose D_c, $x_{(0)}$, $x_{(n)}$ and $[A](D_c)$ are defined as before. We define a restricted sample $\mathcal{R}_n = \{x \in \mathcal{S}_n : x \neq \max \mathcal{S}_n\}$. Let $\mathcal{D}_c$ denote the set of all subsets D_c of the restricted sample $\mathcal{R}_n$, and let $d_{c,n}$ denote the number of elements of $\mathcal{D}_c$, i.e. $d_{c,n} = \#\mathcal{D}_c = \binom{n-1}{c-1}$. Each subset D_c determines a SSP $[A](D_c)$, which further induces a $c \times k$ contingency table $\{N_{ij}(D_c)\}$ $(i = 1, \ldots, c; j = 1, \ldots, k)$ of which the k columns correspond to the k samples. In particular, the counts are given by $N_{ij}(D_c) = \#(A_i(D_c) \cap \mathcal{S}_{j,n_j}) = n_j \int_{A_i} d\hat{F}_{j,n_j}$ $(i = 1, \ldots, c; j = 1, \ldots, k)$.

- *The independence problem.* The size of the SSP is now given by (r, c). Without loss of generality, suppose that $r = \max(r, c)$. We define $D_c^{yz} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{c-1}\} \subset \mathcal{S}_n$ and $D_{r,c}^z = \{\boldsymbol{x}_1', \ldots, \boldsymbol{x}_{r-1}'\}$ such that $D_c^{yz} \subseteq D_{r,c}^z \subset \mathcal{S}_n$. Let $y_{(1)}, \ldots, y_{(c-1)}$ and $z_{(1)}, \ldots, z_{(r-1)}$ denote the ordered Y and Z components of the elements of D_c^{yz} and $D_{r,c}^z$, respectively. Next we define $y_{(0)} = \inf \mathcal{S}_y$, $y_{(c)} = \sup \mathcal{S}_y$, $z_{(0)} = \inf \mathcal{S}_z$ and $z_{(r)} = \sup \mathcal{S}_z$, where $\mathcal{S}_y$ and $\mathcal{S}_z$ denote the (univariate) sample spaces of Y and Z, respectively. An observation-based SSP of size $r \times c$ is then denoted by $[A](D_c^{yz}, D_{r,c}^z)$. For notational comfort we will drop the dependence on the sets D_c^{yz} and $D_{r,c}^z$ where possible. The $r \times c$ SSP is defined as $[A] = \{A_{11}, \ldots, A_{r1}, A_{21}, \ldots, A_{rc}\}$, with elements $A_{kl} = A_k^z \times A_l^y$ $(k = 1, \ldots, r; l = 1, \ldots, c)$, where $A_k^z = \{z \in \mathcal{S}_z : z_{(k-1)} < z \leq z_{(k)}\}$ and $A_l^y = \{y \in \mathcal{S}_y : y_{(l-1)} < y \leq y_{(l)}\}$. Further, we define the restricted samples $\mathcal{R}_n = \{\boldsymbol{x} = (y, z) \in \mathcal{S}_n : y \neq \max \mathcal{S}_n^y \text{ and } z \neq \mathcal{S}_n^z\}$ and $\mathcal{R}_n^z = \{\boldsymbol{x} = (y, z) \in \mathcal{S}_n : z \neq \max \mathcal{S}_n^z\}$. Finally, let $\mathcal{D}_{r,c}$ denote the set of all possible sets D_c^{yz} and $D_{r,c}^z$ containing observations of the restricted sample $\mathcal{R}_n$ and $\mathcal{R}_n^z$, respectively, and let

$$d_{r,c,n} = \#\mathcal{D}_{r,c} = \binom{n}{c-1}\binom{n-(c-1)}{r-c}.$$

Each $r \times c$ SSP $[A](D_c^{yz}, D_{r,c}^z)$ induces an $r \times c$ contingency table $\{N_{kl}(D_c^{yz}, D_{r,c}^z)\}$ where the counts are given by $N_{kl} = \#\,(A_{kl} \cap \mathcal{S}_n) = n \int_{A_{kl}} d\hat{F}_n$.
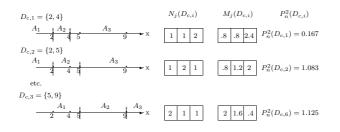

Figure 1. Illustration of the arithmetic involved in the SSPc test. Example: sample $\mathcal{S}_n = \{2, 4, 5, 9\}$; testing for a uniform distribution over $[0, 10]$; SSP size $c = 3$. Thus, $d_{c,n} = d_{3,4} = 6$ different sets $D_{c,i}$ $(i = 1, \ldots, 6)$ (only 3 are shown) and for each of these sets the observed counts $N_j(D_{c,i})$, the expected counts $M_j(D_{c,i})$ $(j = 1, 2, 3)$ and the resulting Pearson statistic $P_n^2(D_{c,i})$ are shown. The SSPc test statistic becomes $T_{c,n} = \frac{1}{6}(0.167 + 1.083 + \ldots + 1.125) = 0.807$.

The Sample Space Partition Test Statistic

The core of the SSP-based test is the Pearson χ^2 goodness-of-fit statistic which is computed on all contingency tables that are induced by the partitions. In particular, in the preceding paragraphs it was explained how d subsets D are derived from the sample $\mathcal{S}_n$ and how each subset D determines a partition $[A](D)$ and how this partition further induces a frequency table. On each such table a Pearson statistic may be computed. The Pearson statistic, denoted by $P_n^2(D)$, is a function of both the observed frequencies and the frequencies expected under the null hypothesis (Table 1).

The SSP test statistic has the general form

$$T_n = \frac{1}{d} \sum_{D \in \mathcal{D}} P_n^2(D). \tag{4}$$

The specific forms of the test statistic for the three goodness-of-fit problems are given in the following paragraphs. Also some properties are given.

- *one-sample problem: the SSPc test.* The test statistic is given by $T_{c,n} = \frac{1}{d_{c,n}} \sum_{D_c \in \mathcal{D}_c} P_n^2(D_c)$. (Figure 1 further illustrates the calculation of the test statistic.) For $c = 2$ its asymptotic null distribution

Table 1. A summary of the observed and the expected frequencies for the 3 goodness-of-fit problems.

	table	obs. freq.	exp. freq
one-sample	$\{N_i\}$	$N_i = n \int_{A_i} d\hat{F}_n$	$M_i = n \int_{A_i} d\hat{G}$
k-sample	$\{N_{ij}\}$	$N_{ij} = n_j \int_{A_i} d\hat{F}_{j,n_j}$	$M_{ij} = n_j \int_{A_i} d\hat{G}$
independence	$\{N_{kl}\}$	$N_{kl} = n \int_{A_{kl}} d\hat{F}_n$	$M_{kl} = n \int_{A_{kl}} d\hat{G}$

has been proven (Thas and Ottoy [18]) for both the simple and the composite null hypothesis. Note that in this simplest case the statistic is closely related to the Anderson-Darling statistic (Anderson and Darling [1]). When $c > 2$ the asymptotic null distribution is conjectured, but approximations are available (Thas [17]).

- *k-sample problem: the SSPkc test.* The test statistic is given by $T_{k,c,n} = \frac{1}{d_{c,n}} \sum_{D_c \in \mathcal{D}_c} P_n^2(D_c)$, which is a rank statistic. When $c = 2$ the test statistic reduces exactly to the k-sample Anderson-Darling statistic of Pettit [8] and Scholtz and Stephens [11]. The SSPkc test may thus be interpreted as an extension of the Anderson-Darling test. When $c > 2$ its exact permutation distribution may be enumerated or approximated by means of Monte Carlo simulation.

- *independence problem: the SSPrc test.* The test statistic is given by $T_{r,c,n} = \frac{1}{d_{r,c,n}} \sum_{(D_c^{xy}, D_{r,c}^z) \in \mathcal{D}_{r,c}} P_n^2(D_c)$, which is again a rank statistic. For $r = c = 2$ the asymptotic null distribution of $T_{r,c,n}$ has been proven (Thas and Ottoy [20]). In this simplest case, the SSPrc test is an Anderson-Darling type test for independence, which, however, has not yet been previously described in the literature. The statistic may also be looked at as Hoeffding's statistic (Hoeffding [5]) with an Anderson-Darling type weight function. Also here, when $c > 2$ the permutation distribution can be enumerated.

Thus the SSPkc and the SSPrc tests are distribution-free rank tests. It has also been proven that all three tests are omnibus consistent for any finite SSP-size. In the next section the power characteristics are briefly summarized.

3.   Power Characteristics

From extensive simulation studies (Thas and Ottoy [18] [19]) it is concluded that the SSPkc and the SSPrc tests are very powerful for many alternatives. In particular, they did not show any power breakdown and their powers were in most of the cases larger than those of many other tests. Even under the few alternatives for which other tests had a larger power, the SSP-based test competed still very well.

Thas [17] performed simulation studies in which the power of the one-sample SSP test is compared to the power of many other tests. In this section we present one of these studies. In particular we test the composite null hypothesis of normality. As an alternative to the normal distribution a family of mixtures of two normal distributions is considered. Its density, indexed by (δ, γ), is given by

$$f_{(\delta,\gamma)}(x) = (1 - \gamma)g(x) + \gamma g \left(\frac{x - \delta}{0.0001} \right),$$

where $g(x)$ is the density of a standard normal distribution. Note that $\gamma = 0$ and $\gamma = 1$ correspond to normal distributions.

The SSP test is compared to five other tests. The best known test is probably the Shapiro-Wilk test (SW) (Shaprio and Wilk [12]), which is implemented as the modified statistic of Weisberg and Binham [23]). Also the Kolmogorov-Smirnov (KS) test is implemented as a modified statistic (Stephens [14] [15]). Since the SSP test closely resembles the Anderson-Darling test (AD), the latter is also included, again as a modified statistic (Stephens [14] [15]). The fourth test is the D'Agostino-Pearson K-test (K) (D'Agostino and Pearson [3]), which is actually only sensitive to deviations in the third and fourth moment. Finally, the Rao-Robson test (RR) (Rao and Robson [9]) is considered. This test is basically a Pearson χ^2 test on equiprobable groups of data. The difference with the SSP based test is that here a grouping is performed only once. We adopted the rule of Mann and Wald [6] to determine the number of groups. In particular, 6 groups were used for $n = 20$ and 9 groups for $n = 50$. We included the SSPc tests with $c = 2, 3$ and $c = 4$.

The critical values at the 5% level of significance are listed in Table 2. The results of the Monte Carlo simulation study are presented in Table 3. All powers were estimated by means of $10,000$ simulation runs. These results suggest that under the particular mixture alternative the SSP tests outperform the other tests, sometimes by a considerable margin. Among the SSP tests the power seems to increase with increasing SSP-size. In power estimation under different alternatives, however, Thas [17] shows

Table 2. Critical Values at the 5%-level.

n	SSP2	SSP3	SSP4	SW	KS	AD	K	RR
20	0.989	4.759	11.037	1.924	0.895	0.752	6.453	10.628
50	0.915	2.813	5.604	2.297	0.985	0.752	6.424	15.206

Table 3. The estimated powers for some alternatives of the contaminated normal family, based on 10000 simulation runs.

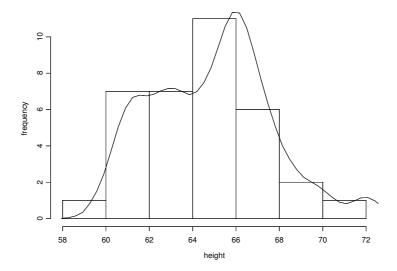| n	δ	γ	power							
			SSP2	SSP3	SSP4	SW	KS	AD	K	RR
20	0	0.2	0.215	0.769	0.781	0.223	0.325	0.261	0.133	0.270
20	0	0.7	1.000	1.000	1.000	0.997	1.000	1.000	0.823	1.000
20	2	0.4	0.683	0.991	0.991	0.797	0.810	0.857	0.230	0.845
20	4	0.4	0.948	0.996	0.996	0.972	0.949	0.990	0.624	0.881
50	0	0.2	0.513	0.986	0.990	0.415	0.706	0.580	0.173	0.757
50	0	0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.875	1.000
50	1.25	0.38	0.993	1.000	1.000	0.990	0.993	0.997	0.541	0.999
50	3	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.938	0.999
50	3.5	0.1	0.754	0.821	0.804	0.844	0.550	0.748	0.431	0.470s

that the power may just as well decrease with increasing SSP-size. This clearly illustrates the importance of choosing a good size. In practice, though, the user most often does not know what the optimal SSP-size is in advance. As a solution the authors have proposed data-driven SSP tests (Thas and Ottoy [21]), which avoid the problem by estimating the SSP-size from the data by utilising a selection rule.

Although the results presented in Table 3 may seem very convincing, simulations under other alternatives have indicated that sometimes the power of the SSP test is considerably lower when compared to the other tests (Thas [17]). This is unfortunately a characteristic of all omnibus consistent tests.

4. Example of the One-Sample SSP Test

The Singer dataset, which was used by Cleveland [2] to demonstrate Trellis graphs, consists of heights of singers in the New-York Choral Society. Here, we only consider the group of 35 alto. The Shapiro-Wilk test and the Kolmogorov-Smirnov test resulted in $p = 0.379$ and $p = 0.309$, respectively. Since none of the p-values is small, in conclusion, there does not seem to be much evidence against the hypothesis of normality.

The computed test statistics (p-values) obtained with the SSPc tests with $c = 2, 3$ and $c = 4$ are $t_2 = 0.452$ ($p = 0.374$), $t_3 = 306.883$ ($p < 0.0001$) and $t_4 = 951.133$ ($p < 0.0001$), respectively. The results of the SSP2 test agrees quiet well with the two classical tests, but when larger partition sizes are used, one immediately notices the extreme large values of the test statistics which correspond to p-values smaller than 0.0001. Thus, both the SSP3 and SSP4 test reject the null hypothesis of normality. A reason for this tremendous difference between the SSP2 test, on the one hand, and the SSP3 and SSP4 tests on the other hand, might be that the data actually shows bimodality. Simulation results have indeed pointed out that extremely high powers of the SSP3 and SSP4 tests are especially observed under bimodal alternatives. Figure 2 shows the histogram of the data and a Gaussian kernel density estimate with a bandwidth manually set to 2.75, suggesting a bimodal distribution. However using the Unbiased Cross Validation bandwidth selection method (Silvermann [13]; Venables and Ripley [22]), the data would have been oversmoothed in the sense that the bumps in the density are all flattened, which is a typical characteristic of such a method in small samples.

Figure 2. A histogram and a Gaussian kernel density estimate (bandwidth = 2.75) of the Singer Data .

References

1. T. Anderson and D. Darling. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 193–212, 1952.
2. W. Cleveland. *Visualizing Data*. New Jersey: Hobart Press, 1993.
3. R. D'Agostino and E. Pearson. Testing for departure from normality. I. Fuller empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika* 60:613–622, 1973.
4. R. Fisher. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1925.
5. W. Hoeffding. A non-parametric test of independence. *Annals of Mathematical Statistics* 19:546–557, 1948.
6. H. Mann and A. Wald. On the choice of the number of class intervals in the application of the chi-square test. *Annals of Mathematical Statistics* 13:306–317, 1948.
7. D. Moore. *Tests of chi-squared type*. In: R. D'Agostino and M. Stephens (Eds.), Goodness-of-fit Techniques. Marcel Decker, 1986.
8. A.Pettit. A two-sample Anderson-Darling rank statistic. *Biometrika* 63:161–168, 1976.
9. K. Rao and D. Robson. A chi-square statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics* 3:1139–1153, 1974.
10. J. Rayner and D. Best. *A Contingency Table Approach to Nonparametric Testing*. Boca Raton: Chapman and Hall, 2001.
11. F. Scholtz and M. Stephens. k-sample Anderson-Darling tests. *Journal of the American Statistical Association* 82:918–924, 1987.
12. S. Shapiro and M. Wilk. An analysis of variance test for normality complete samples. *Biometrika* 52:591–611, 1965.
13. B. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986.
14. M. Stephens. EDF statistics for goodness-of-fit and some comparisons. *Journal of the American Statistical Association* 69:730–737, 1974.
15. M. Stephens. Asymptotic results for goodness-of-fit statistics with unknown parameters. *Annals of Statistics* 4:357–369, 1976.
16. M. Stephens. Tests based on EDF statistics. In: R. D'Agostino and M. Stephens (Eds.). *Goodness-of-fit Techniques*, 97–193. New York: Marcel Dekkers, 1986.
17. O. Thas. Nonparametric Tests Based on Sample Space Partitions. PhD thesis. *Ghent University, Belgium*, 2001.
18. O. Thas and J. P. Ottoy. An extension of the Anderson-Darling k-sample test to arbitrary sample space partition sizes. Submitted, 2002.
19. O. Thas and J. P. Ottoy. A generalization and an extension of the Anderson-Darling statistic. Submitted, 2002.
20. O. Thas and J. P. Ottoy. A nonparametric test for independence based on sample space partitions. Submitted, 2002.
21. O. Thas and J. P. Ottoy. Data-driven versions of the SSPc, SSPkc and SSPrc tests. Submitted, 2002.
22. W. Venables and B. Ripley. *Modern Applied Statistics with S-plus*. Springer, 1997.
23. S. Weisberg and C. Binham. An approximation analysis of variance test for non-normality suitable for machine calculation. *Technometrics* 17:133–134, 1975.