HEN-419
L3 Note 2356
January 25, 1999

The Minimum Spanning Tree as a Two-Sample Test

W. J. Metzger

Experimental High Energy Physics Group
University of Nijmegen
Nijmegen, The Netherlands

1 Introduction

Given two independent samples of observations, we may ask the question whether both samples are derived from the same distribution, *i.e.*, the same probability density function (p.d.f.). If we are content to bin the data, we can use a χ^2 test. However, this assumes a 'large' number of observations per bin, and the limited number of observations may necessitate using bins larger than the experimental resolution, with a resulting loss of information.

If the observations are in one dimension there are several well-known standard tests, *e.g.*, the Smirnov-Cramér-von Mises test and the Kolmogorov (or Smirnov, or Kolmogorov-Smirnov) test, both of which make use of the cumulative distribution of the points. There is also the run test. Descriptions of these tests can be found in most statistics books.

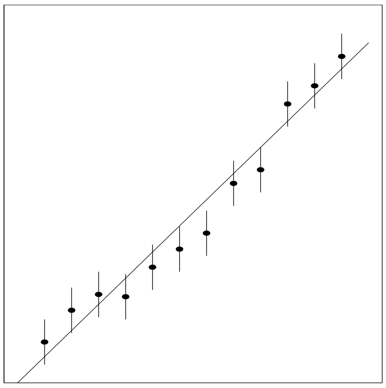
However, in more than one dimension, it is not so easy. One can always bin and use χ^2, but the disadvantages of binning usually become much greater as the dimensionality increases.

The Smirnov[1] (or Kolmogorov, or Kolmogorov-Smirnov) test has been generalized to two dimensions,[2] and routines are given in *Numerical Recipes*.[3] Although the significance level of the test is only approximate, it can still provide useful discrimination.

In this note we present tests based on the minimum spanning tree. These tests are generalizations of the one-dimensional run test and the Kolmogorov-Smirnov test. The tests are not new, having been proposed 30 years ago,[4] but are little known.

2 Run test

The run test[5] is usually introduced as a goodness-of-fit test comparing a set of points to a predicted function. The χ^2 test makes use of the squares of the deviations of the data from that expected under H_0. Thus it only uses the magnitudes of the deviations and ignores their signs. However, the signs of the deviations are also important, and systematic deviations of the same sign indicate that the hypothesis is unlikely, as is illustrated in the figure.



A test which uses only the sign of the deviations is the run test. A run is defined as a set of adjacent points all having the same sign of deviation. The data and curve in the figure have deviations $AAABBBBBBAAA$, where A represents a positive and B a negative deviation. There are thus three runs, which seems rather small. We would expect the chance of an A to equal that of a B and to show no correlation between points if the hypothesis were true. This implies that we should expect runs to be short; a long run of 6 points as in the figure should be unlikely. In fact, this expectation is strictly true only if H_0 is a simple hypothesis.

To be more quantitative, let k_A be the number of positive deviations and k_B the number of negative deviations observed. Let $k = k_A + k_B$. Given k_A and k_B, we can calculate the probability that there will be r runs. If either k_A or k_B is zero, there is, necessarily only one run, and $P(r = 1) = 1$.

Given k_A and k_B, the number of different ways to arrange them is

$$\binom{k}{k_A} = \frac{k!}{k_A!k_B!}$$

Suppose that there are r runs. First, suppose that r is even and that the sequence begins with an A. Then there are k_A A-points and $r/2 - 1$ divisions between them. For the example of the figure this is $AAA|AAA$. With k_A A's there are $k_A - 1$ places to put the first dividing line, since it can not go at the ends. Then there are $k_A - 2$ places to put the second dividing line, since it can not go at the ends or next to the first dividing line. In total there are $\binom{k_A-1}{r/2-1}$ ways to arrange the dividing lines among the A's. There is a similar factor for arrangement of the B's and a factor 2 because we assumed we started with an A and it could just have well been a B. Thus the probability of r runs, for r even, is

$$P(r) = \frac{2\binom{k_A-1}{r/2-1}\binom{k_B-1}{r/2-1}}{\binom{k}{k_A}} \tag{1}$$

Similarly, one finds for r odd

$$P(r) = \frac{\binom{k_A-1}{(r-3)/2}\binom{k_B-1}{(r-1)/2} + \binom{k_A-1}{(r-1)/2}\binom{k_B-1}{(r-3)/2}}{\binom{k}{k_A}} \tag{2}$$

From these it can be shown that the expectation and variance of r are

$$E[r] = 1 + \frac{2k_Ak_B}{k} \tag{3}$$

$$V[r] = \frac{2k_Ak_B(2k_Ak_B - k)}{k^2(k-1)} \tag{4}$$

The critical region of the test is defined as improbably low values of r, $r < r_\alpha$. For k_A and k_B greater than about 10 or 15, one can use the Gaussian approximation for r. For smaller numbers one can compute the probabilities directly using equations 1 and 2. In our example, $k_A = k_B = 6$. From equations 3 and 4 we expect $r = 7$ with variance 2.73, or $\sigma = 1.65$. We observe 3 runs, which differs from the expected number by $4/1.65 = 2.4$ standard deviations. Using the Gaussian approximation, this corresponds to a (one-tailed) confidence level of 0.8%. Exact calculation using equations 1 and 2 yields $P(1) + P(2) + P(3) = 1.5\%$. Whereas the χ^2 is acceptable ($\chi^2 = 12$ for 12 points), the run test suggests that the curve does not fit the data.

The run test is much less powerful than a χ^2 test, using as it does much less information. But the two tests are completely independent and hence they can be combined. An hypothesis may have an acceptable χ^2, but still be wrong and rejectable by the run test. Unfortunately, the run test is applicable only when H_0 is simple. If parameters have been estimated from the data, the distribution of the number of runs is not known and the test can not be applied.

To test whether two samples are from the same p.d.f., one can use the same run test, except that runs are defined as runs in sample number rather than in the sign of the deviation from a prediction. Thus the combined sample is sorted in the variable and the sample numbers are noted as one progresses from lowest to highest, *e.g.*, $AAABBBBBBAAA$, where A and B represent sample 1 or 2, respectively. The situation is the same as above: the number of runs is three, and the confidence level is computed as before. Note now that instead of the number of runs we could count the number of times, t, a transition occurs from one sample to another (A to B or B to A). Clearly, $t = r - 1$.

3 Kolmogorov-Smirnov test

This test is based on the maximum difference in the cumulative distribution functions (c.d.f.) of the two samples:

$$D_{n_1 n_2} = \max\{|S_{n_1}(x) - S_{n_2}(x)|\} \qquad \text{for all } x \tag{5}$$

where the sample c.d.f., $S_i(x)$, is the fraction of points below x in sample i. If $D_{n_1 n_2}$ is large, it is unlikely that the samples are the same. The distribution of $\sqrt{\frac{n_1 n_2}{n_1+n_2}} D_{n_1 n_2}$ is known and from it the confidence level can be calculated.

4 Minimum spanning tree

The run test and the Kolmogorov-Smirnov test are illustrated in Fig. 1. It should be apparent that both tests only make use of the order (or rank) of the points in x, not of the values of x themselves.

The problem in generalizing the run test (or the Smirnov test) to more than one dimension is that there is no unique prescription for sorting the points. The minimum spanning tree (MST) can be used for this purpose.[4]

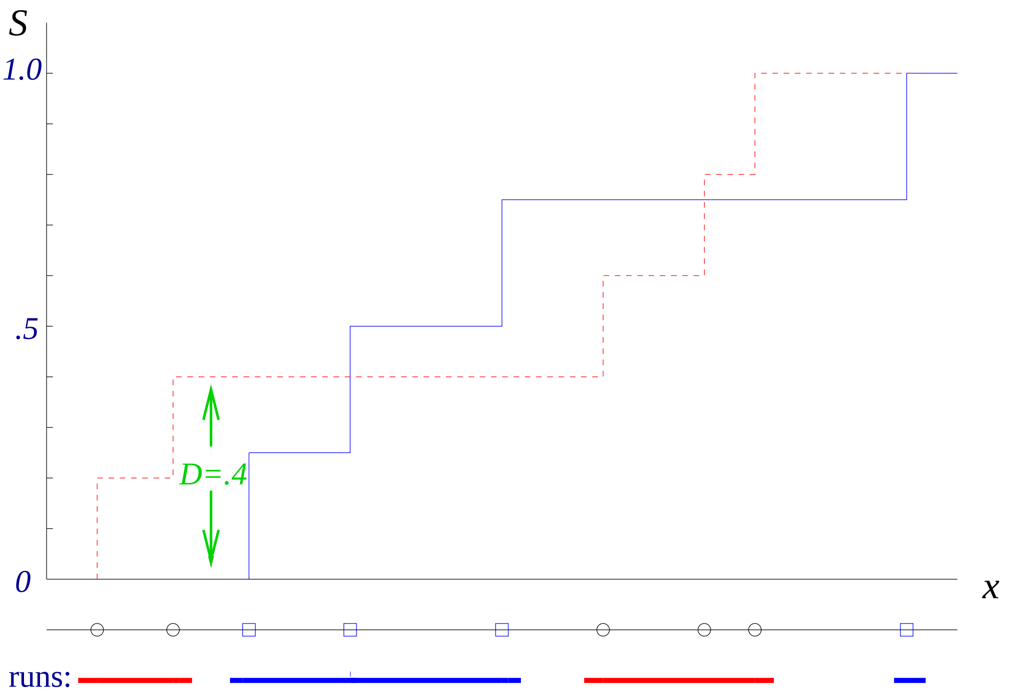


Figure 1: Illustration of the Kolmogorov-Smirnov and the run tests in one dimension. Two samples of points (5 circles and 4 squares) result in 4 runs corresponding to a CL of 26%. The sample c.d.f., S of each sample is shown; the maximum difference is $D = 0.4$, giving a K-S test statistic of $\sqrt{\frac{5\cdot 4}{5+4}}D = 0.27$, corresponding to a CL of 87%.

A spanning tree is a set of discrete points all of which are connected such that the connections (called **edges**) have no loops. Or, in other words, starting from any point, i, it is impossible to return to that point in any way except by retracing the path you have taken. A minimum spanning tree is the spanning tree for which the total length of the connections is the smallest possible. In principle the minimum spanning tree is not necessarily unique; there may be more than one spanning tree with the same minimal length if there are two or more edges of identical length. However, in practice we probably do not have to worry about that. In one dimension the MST is trivial and corresponds to simply sorting the points.

5 Two-sample MST Tests

The tests discussed in this section have all been discussed (in more detail) elsewhere.[4] We will only treat them in the 'asymptotic' limit, which fortunately is reached quite quickly (say 10 or 20 points for the run test and 80 or so for Smirnov).

5.1 Generalized run test

To use the minimum spanning tree as a test of whether two samples are from the same p.d.f., we construct an MST using all the points of both samples. We then count the number of connections between points from different samples, which are called **incompatible edges**. This corresponds to the number of transitions, t, in section 2.

The probability of t incompatible edges is the same as the probability of $r = t+1$ runs (equation 3), except for an additional covariance term. If we assign each edge i a value $z_i = 0$ or 1, for compatible and incompatible edges, respectively, the number of runs and its expectaion are given for k points by

$$r = 1 + \sum_{i=1}^{k-1} z_i \qquad \text{and} \qquad E\left[r\right] = 1 + \sum_{i=1}^{k-1} E\left[z_i\right] \tag{6}$$

The expectation of z_i is just the probability that $z_i = 1$ which is equal to the fraction of edges whose endpoints are from different samples, AB or BA. This is given by

$$P\{AB\} = P\{BA\} = \frac{k_A k_B}{k(k-1)} \tag{7}$$

Adding $P\{AB\}$ and $P\{BA\}$ we get the same result as in equation 3. However, the variance of the distribution of number of runs is more complicated than equation 4. This comes about because of the covariance between incompatible edges.

The variance of the number of runs is

$$V\left[r\right] = V\left[\sum_{i=1}^{k-1} z_i\right] \tag{8}$$

$$= \sum_{i=1}^{k-1} V\left[z_i\right] + 2\sum_{i<j} \operatorname{cov} z_i z_j \tag{9}$$

Now,

$$\operatorname{cov} z_i z_j = E\left[z_i z_j\right] - \left(E\left[z_i\right]\right)^2 \tag{10}$$

The difficulty comes in evaluating $E\left[z_i z_j\right]$. $z_i z_j = 1$ only if edges i and j are both incompatible, which depends on whether they have a common endpoint. If they do, the possibilities are ABA or BAB. However, if not, the edges can be AB, AB, AB, BA, BA, BA, or BA, AB, which leads to a different expression. It can be shown[4] that the variance is given by

$$V\left[r\right] = \frac{2k_A k_B}{k(k-1)} \left\{ \frac{2k_A k_B - k}{k} + \frac{C - k + 2}{(k-2)(k-3)} \left[k(k-1) - 4k_A k_B + 2\right] \right\} \tag{11}$$

where C is given by

$$C = \frac{1}{2} \sum_{i=1}^{k} d_i (d_i - 1) \tag{12}$$

with d_i the number of edges with point i as an endpoint. Note that in the one-dimensional case $d_i = 2$ for all except 2 points (the smallest and the largest), for which $d_i = 1$. Hence, $C = k - 2$ and equation 11 reduces to equation 4. You might think that C would have to be taken from the parent distribution, which you generally do not know. Further, the test would then not be "distribution-free". By taking C from the data, both of these problems are overcome.

The confidence level of the test is the probability of t or fewer incompatible edges, and may be easily computed in the asymptotic limit where r is distributed normally.

5.2 Additional MSTs

For a sample of k points, the number of edges in the MST is $k - 1$. However, the number of pairs of points is $k(k-1)/2$. Thus, for increasing sample size, there are many pairs of points which are close together but are not MST edges. Since the power of a test will usually depend on recognizing local differences, it would be beneficial to include more pairs of points which are close together. This can be done by constructing a second MST which is orthogonal to the first, *i.e.*, having no edges in common. One can continue the process to construct additional orthogonal MSTs, although the gain in power is expected to decrease. Clearly, the number of MSTs used should be much smaller than the number of points.

Now, the results of section 5.1 did not at all depend on the fact that we used an MST. Any tree of $k - 1$ edges would do. The MST was used because of its property of mainly connecting points which are close together. The formalism can be immediately applied to the $n(k-1)$ edges of the n orthogonal MSTs.

5.3 Degree-1 nodes

The degree of a point (or node) is defined as the number of edges connected to the point (called d_i in the section 5.1). Degree-1 nodes occur mainly at the edges of the sample. If the two samples are the same, each will have about the same number of degree-1 nodes. Unfortunately, this is also the case if the two samples are completely separated, *e.g.*, two Gaussians whose means differ by much more than their standard deviations. However, *e.g.*, two Gaussians with the same mean but different standard deviations will tend to have degree-1 nodes predominantly in the sample of the wider Gaussian. A test can be formulated based on the expected number of degree-1 nodes in the two samples.[4]

5.4 Smirnov

Based, as it is on the c.d.f., this test requires the points to be ordered. Ordering in one dimension is generally obvious (though not always so, *e.g.*, using x or x^2 when x can be both positive and negative) and in any case straightforward. Ordering in more than one dimension is not. The MST can be used as the ordering principle. Just as in the one-dimensional case, adjacent points are close together. But besides deciding to use the MST to order, one must also decide where to start. In one dimension one starts at the minimum (or the maximum) value of x, which is also the value of x furthest from the middle value.

Let us define the **eccentricity** of a node: Consider all paths from a point i to an endpoint. The number of edges in the path having the most edges is called the eccentricity of the point. The MST analog of the Smirnov test takes as starting point one of the points having the largest eccentricity. This is clearly an endpoint, and there are at least two such points. The analog of the c.d.f. of a sample is then the number points from that sample as function of the number of edges from the starting point. At a branch of the tree one follows the shorter branch (in terms of number of edges) completely to its end before proceeding with the longer branch. In the case of branches with the same number of edges, one takes the branch with the shortest real distance first.

Just as with the one-dimensional case, this test is not very powerful when the difference in the two samples is mainly one of scale, rather than location of the mean. This can be cured by ranking in distance from the median of the combined sample, at the expense of power with respect to difference in location. The MST analog of this procedure, called "radial" Smirnov, is to take as starting point the node of minimal eccentricity, which in one dimension is the same as the median.

An example of an MST, which illustrates these tests, is shown in Fig. 2. The sample c.d.f.s for the standard Smirnov test on this MST are shown in Fig. 3.

6 FORTRAN routines

To construct the tree, all that is needed is a matrix with the distance between each pair of points. I have written two routines, one to construct the MST* and the other to calculate the confidence level for the MST generalization of the run test. Note that they require the points to be ordered by sample number. The routines are used as follows:

```
!     Integer NA, NB  = number of points in sample A and B
!     Integer NP = NA+NB
!     Real    D(NP*(NP-1)/2) = the (triangular) distance array
!     Integer I2J(2:NP)
!     Real    D2J(2:NP)
!     Integer DEG(NP)
!
!     calculate distance matrix
!     This assumes the first NA points are from sample A
!     and that the remaining points are from sample B
      dlarge = 0.
      nd = 0
      DO j = 2, NA + NB
```

*The routine to construct the MST is based on an an ALGOL routine[6] which incorportates an algorithm by Prim.[7] This algorithm is very efficient for a few hundred data points. For much larger data sets, other algorithms become more efficient

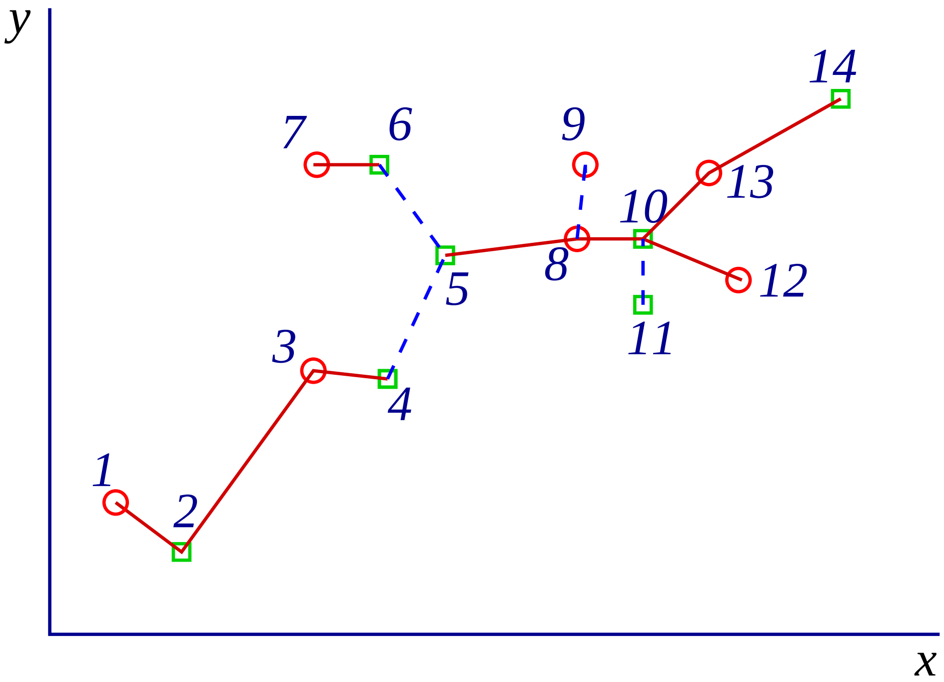


Figure 2: Illustration of an MST of two samples of points (7 circles and 7 squares). There are 9 incompatible and 4 compatible edges. The maximum eccentricity, 8, is for points 1 and 14. The numbering of the points corresponds to the order for the standard Smirnov test. Point 5 has the smallest eccentricity, 4. The radial Smirnov test's order is 5, 6, 7, 4, 3, 2, 1, 8, 9, 10, 11, 12, 13, 14. There are 6 degree-1 nodes.

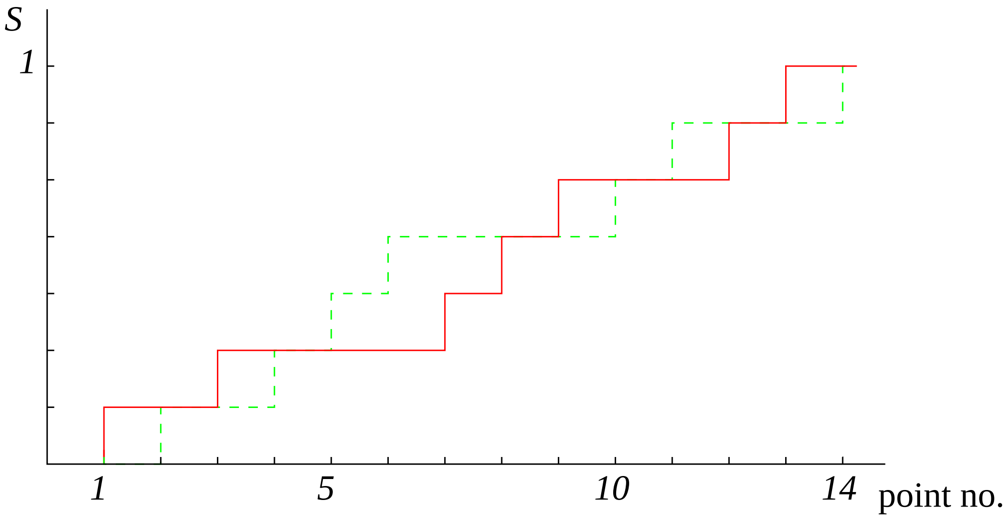


Figure 3: The sample c.d.f.s for the standard Smirnov test for the MST of Fig. 2. The maximum difference of 2/7 occurs between points 6 and 7.

```
         DO i = 1, j - 1
            nd = nd + 1
            D(nd) = distance between points i and j
            dlarge = MAX(dlarge,D(nd))
         ENDDO
      ENDDO
*
*     construct tree
!     on return from MSTREE,
!         I2J(i) is the endpoint of an edge beginning at point i
!         D2J(i) is the length of the edge described by I2J(i)
      CALL MSTREE(NA+NB,LOGi,dlarge,D,I2J,D2J,ifault)
      IF ( ifault.NE.0 ) THEN
         PRINT *, 'MSTREE error', ifault
         STOP
      ENDIF
*
*     calculate confidence level
!     the array DEG is used as temporary memory in CLMST
      cl = CLMST(NA,NB,I2J,DEG)
```

Note that although `D` is called the distance matrix, any monotonically increasing function of `D` will result in the same MST, and hence the same CL, since only comparisons of distances between pairs of points are used in constructing the tree, not the distances themselves. (In particular, the square of the distance results in the same MST as the distance.)

Alternatively, you can use routines by J. Friedman, which also calculate the other MST confidence levels mentioned in the previous section. I have modified his routines somewhat to give control of the distance calculation to the user rather than the hard-coded Euclidean distance of the original routines.†

```
!     NA,NB,dlarge,D are the same as above
!     NTREE is the number of trees to include in the runs tests
!         The above test (CLMST) is for just 1 tree.
!     Real    DEG(NP,2)               -- temporary array for MSTCLFRD
!     Integer MST(max(5,NTREE+3),NP) -- temporary array for MSTCLFRD
!     Real CLs(NTREE+3)  the resulting confidence levels:
!          CLs(1) = conf. level from degree-1 test
!          CLs(2) = conf. level from Smirnov test (powerful for location)
!          CLs(3) = conf. level from Smirnov test (powerful for scale)
!          CLs(4) = conf. level from run test - 1 MST (same as CLMST)
```

†I have also changed some names and calling sequences. I have replaced the routine which calculates the CL for the Kolmogorov-Smirnov tests by the one from *Numerical Recipes*,[3] and removed his routine which calculates the error function, since it (`ERF`) is available on most computers these days. (If it is not available in the standard libraries, it can be found in the CERN libraries.)

```
!          CLs(5) = conf. level from run test - 2 MSTs
!          ...
!          CLs(NTREE+3) = conf. level from run test - NTREE MSTs
!
!
!     Calculate distance matrix as above, and then use MSTCLFRD
!     to both construct the MST and calculate the confidence levels
!
      CALL MSTCLFRD(NA,NB,DLARGE,D,NTREE,deg,mst,CLs)
```

The routines `MSTREE`, `CLMST`, and `MSTCLFRD` are available at `http://home.cern.ch/~metzger`.

7 Test of the method

To test the routines and the method, I generated samples of 100 Monte Carlo points using a two-dimensional normal p.d.f. with zero correlation. Each dimension is a normal distribution with the same mean and standard deviation. The samples generated have

1. mean = 0, standard deviation = 1; or
2. mean = 0.5, standard deviation = 1; or
3. mean = 0, standard deviation = 1.5.

Confidence levels (CL) are calculated for two type-1 samples, a type-1 sample with a type-2 sample and a type-1 sample with a type-3 sample. This was repeated 1000 times. The resulting CL distributions are shown in the figures.

Figure 4 shows the results for two type-1 samples. The first plot is a histogram of the number of incompatible edges. The next shows the confidence level calculated from the number of incompatible edges. This plot of the confidence level contains two histograms. One is a series of discrete values, which occurs since the number of incompatible edges is an integer. For a better display, each discrete value is converted to a continuous value by generating a random number uniformly distributed from the mid-point between the value and the next lower value and the mid-point between the value and the next higher value. As expected for two samples which are from the same p.d.f., this distribution is flat. The mean and rms (whether from the continuous or from the discrete distribution are consistent with that of a uniform distribution, namely mean of 0.5 and rms of $\sqrt{1/12} = 0.2887$. It was checked that both `MSTREE/CLMST` and `MSTCLFRD` give the same results.

Figure 4 also shows the confidence level for the two-dimensional Kolmogorov-Smirnov test from *Numerical Recipes*[3] (KS-NR). It is apparent that the confidence level is not uniformly distributed, reflecting the approximate nature of this test. Further, we see in the scatter plot of the KS-NR CL *vs.* the MST CL that the two tests are only slightly correlated. This is not surprising, since the KS-NR test is based on a maximum deviation in the cumulative distributions of the two samples whereas the MST is based on an ordering of the points.

Figure 5 shows the confidence levels from the "standard" and "radial" Smirnov tests based on the MST. Both distributions of confidence levels have a mean somewhat less than 0.5, but the rms is quite close to that expected for a uniform distribution. In any case these distributions are much closer to a uniform distribution than that of the Smirnov test from *Numerical Recipes*. Also shown are the confidence level distributions for the test using the number of degree-1 nodes and from using additional (orthogonal) MSTs, all of which are consistent with uniform. Figure 6 shows the correlations between various MST confidence levels.

Next type-1 distributions are compared with type-2 (Figs. 7 and 8) and with type-3 (Figs. 9 and 10) distributions. In both cases the distribution of incompatible edges shifts to lower values and the confidence levels become peaked at small values. It appears that the KS-NR CL is more peaked at low values than the MST CL, particularly when the distributions differ in their mean. This could indicate that KS-NR provides a more powerful test. However, one must be very cautious about this since this KS test is only approximate; it was already skewed to low values when samples from the same p.d.f. were compared. The power of the KS tests based on the MST depends on the distributions. As expected, the standard KS test is powerful when the distributions differ in the mean (Fig. 8) while the radial KS test is powerful when the difference is in the width (Fig. 10). The degree-1 test is seen to have no power when the difference is in the mean and is clearly less powerful than the radial KS or the MST run tests when the difference is in the width. Using additional, orthogonal, MSTs increases the power of the MST run test, as expected, and approaches the power of the best KS test when enough MSTs are used. The power of the MST run test is much less dependent on the differenece between the distributions (mean or width) than are the other tests.

A great advantage of the MST tests is the triviality of extending them to more dimensions. The only thing that changes is the calculation of the distance for the distance matrix. The tests themselves do not use the dimensionality, only the distances and the numbers of points in each sample. As an example we show in Figs. 11–14 MST results for sets of 100 points in 4 dimensions. Again we generate multi-dimensional normal distributions, with no correlations and the same mean and standard deviation for each dimension. The three types are as before, but with means 0 or 0.3 and standard deviations 1.0 or 1.3. Results from comparing two type-1 samples are shown in Fig. 11. The CLs are uniform, as expected. Comparisons of type-1 and type-2 samples and of type-1 and type-3 samples are shown in Figs. 13 and 14. The analogous plots for 16 dimensions are shown in Figs. 15-18. The behavior of the CLs in higher dimensions is similar to the 2-dimensional case.

To get an idea of the dependence of the relative power of the various tests on the dimensionality, Tables 7 and 7 shows the number of trials (out of 1000) having a CL less than 1%. For small dimensionality, the appropriate Smirnov test is more powerful the others. However this difference decreases as the dimensionality increases. The degree-1 nodes test has no power in the examples of differing means, but is very powerful, particularly at high dimensionality in the example of differing

	2-dim.	4-dim.	8-dim	16-dim
Std. Smirnov	630	195	168	200
Rad. Smirnov	145	78	74	94
Degree-1	5	4	3	4
1-MST run	212	83	242	461
2-MST run	368	169	418	720
4-MST run	555	321	610	890

Table 1: The number of trials (out of 1000) with CL less than 1% when comparing two multi-dimensional normal distributions differing in means by 0.3 (0.5 in the 2-dim. case) but both with standard deviations 1.

	2-dim.	4-dim.	8-dim	16-dim
Std. Smirnov	69	19	14	7
Rad. Smirnov	652	375	508	720
Degree-1	45	242	855	999
1-MST run	217	116	149	199
2-MST run	335	200	320	440
4-MST run	555	297	490	740

Table 2: The number of trials (out of 1000) with CL less than 1% when comparing two multi-dimensional normal distributions with the same means but with standard deviations 1 and 1.3 (1 and 1.5 in the 2-dim. case).

widths. The relative power of the MST run tests increases with dimensionality and does not depend so much on whether it is the mean or width which differs between the samples.

8 Conclusions

The minimum spanning tree provides methods to compute a confidence level for the hypothesis that two samples are derived from the same distribution. The test is applicable for any number of dimensions. The power of the Kolmogorov-Smirnov tests and of the test based on the number of degree-1 nodes depends strongly on how the distributions differ. For low dimensionality one of these tests may have greater power than the MST run test. However, as the dimensionality increases this difference becomes less. The power of the MST run test seems less dependent on how the distributions differ.

Acknowledgments

Thanks to Ljubisa Drndarski and Jaap Schotanus for pointing out the possibility of using the MST as a test. Thanks to Thomas Hebbeker for calling my attention to the two-dimensional K-S test in *Numerical Recipes*. I am grateful to J. Friedman for giving me his routines, which despite being written 20 or more years ago still work well.

References

[1] N. V. Smirnov, *Bull. Moscow Univ.* **2** (1987) 3.

[2] G. Fasano and A. Franceschini, *Monthly Notices of the Royal Astron. Soc.* **225** (1987) 155.

[3] William H. Press *et al.*, *Numerical Recipes in FORTRAN77: The Art of Scientific Computing*, Cambridge Univ. Press.

[4] Jerome H. Friedman and Lawrence C. Rafsky, *Ann. Statist.* **18** (1969) 103.

[5] A. Wald and J. Wolfowitz, *Ann. Math. Stat.* **11** (1940) 147-162.

[6] G. J. S. Ross, *Appl. Statist.* **18** (1969) 103.

[7] R. C. Prim, *Bell System Tech. J.* **36** (1957) 1389.

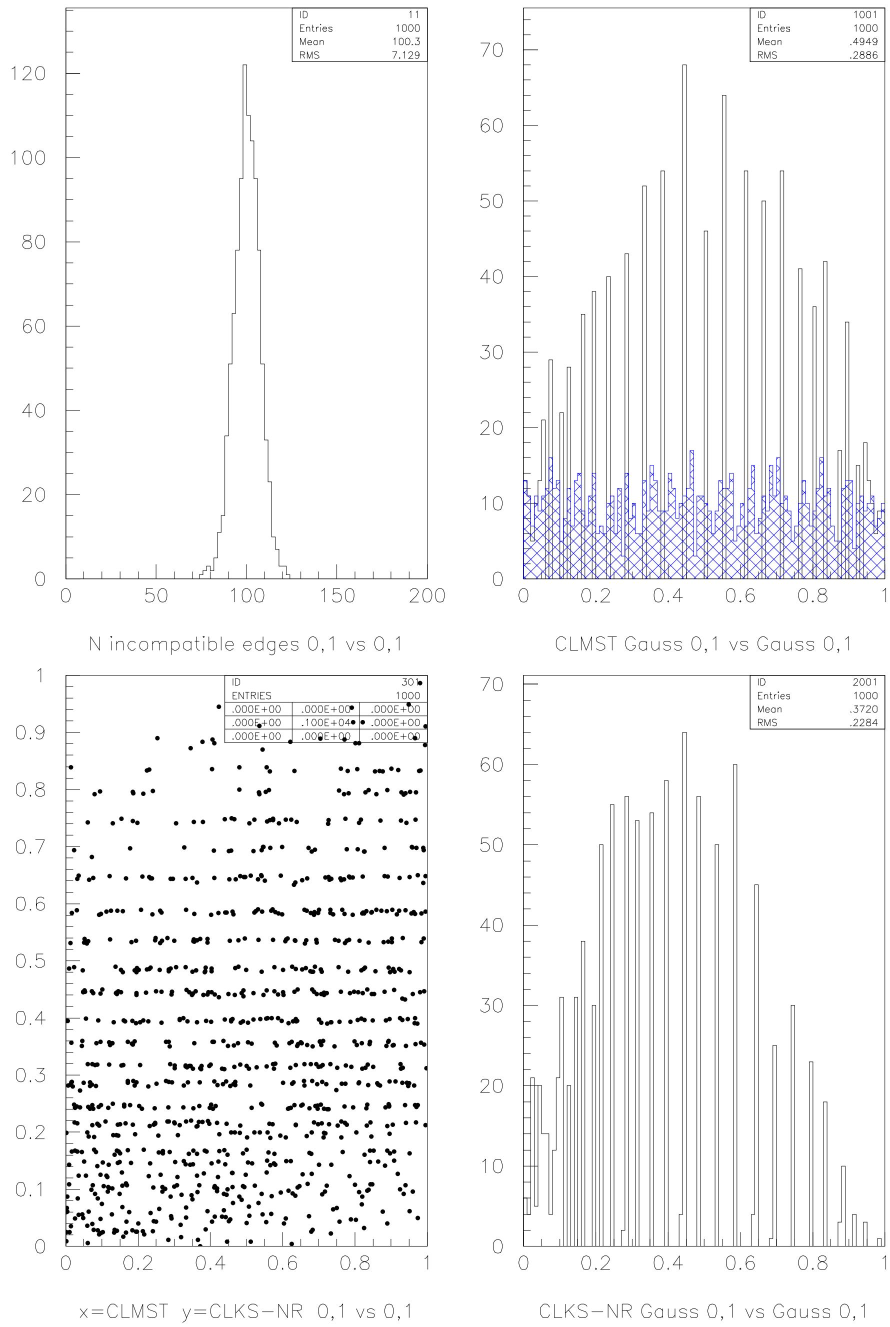


Figure 4: Comparison of 2 samples of 100 points, both from a 2-dimensional standard normal distribution, for 1000 pairs of samples: number of incompatible edges in the MST; confidence level calculated from the number of incompatible edges (the cross-hatched histogram is obtained from the discrete histogram by randomizing the discrete values); scatter plot of this MST confidence level and that from the 2-dimensional Kolmogorov-Smirnov test from *Numerical Recipes*; and confidence level of this 2-dimensional Kolmogorov-Smirnov test.

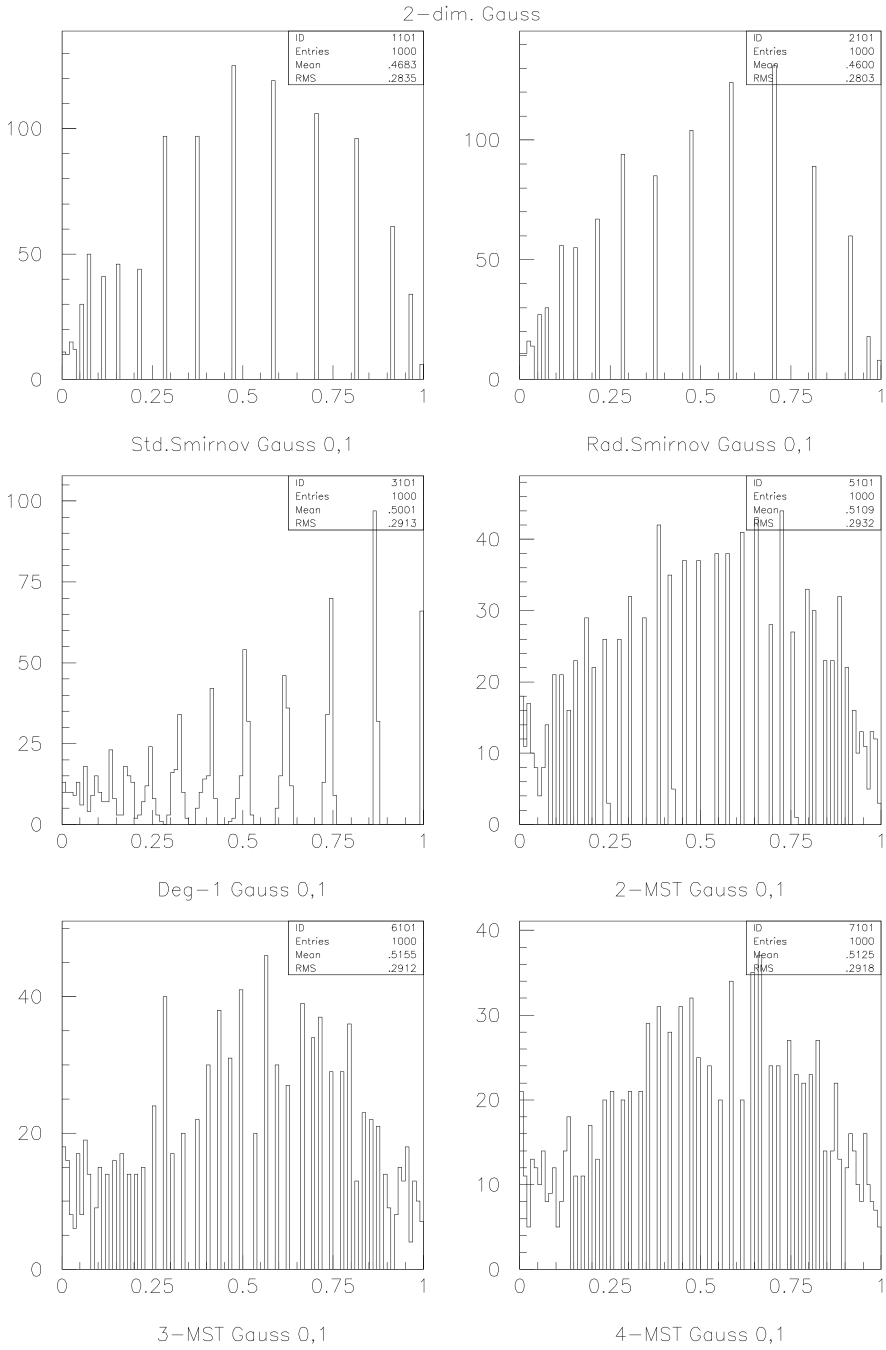


Figure 5: Confidence levels for 2 samples of 100 points, both from a 2-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

Figure 6: Correlations between confidence levels for 2 samples of 100 points, both from a 2-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

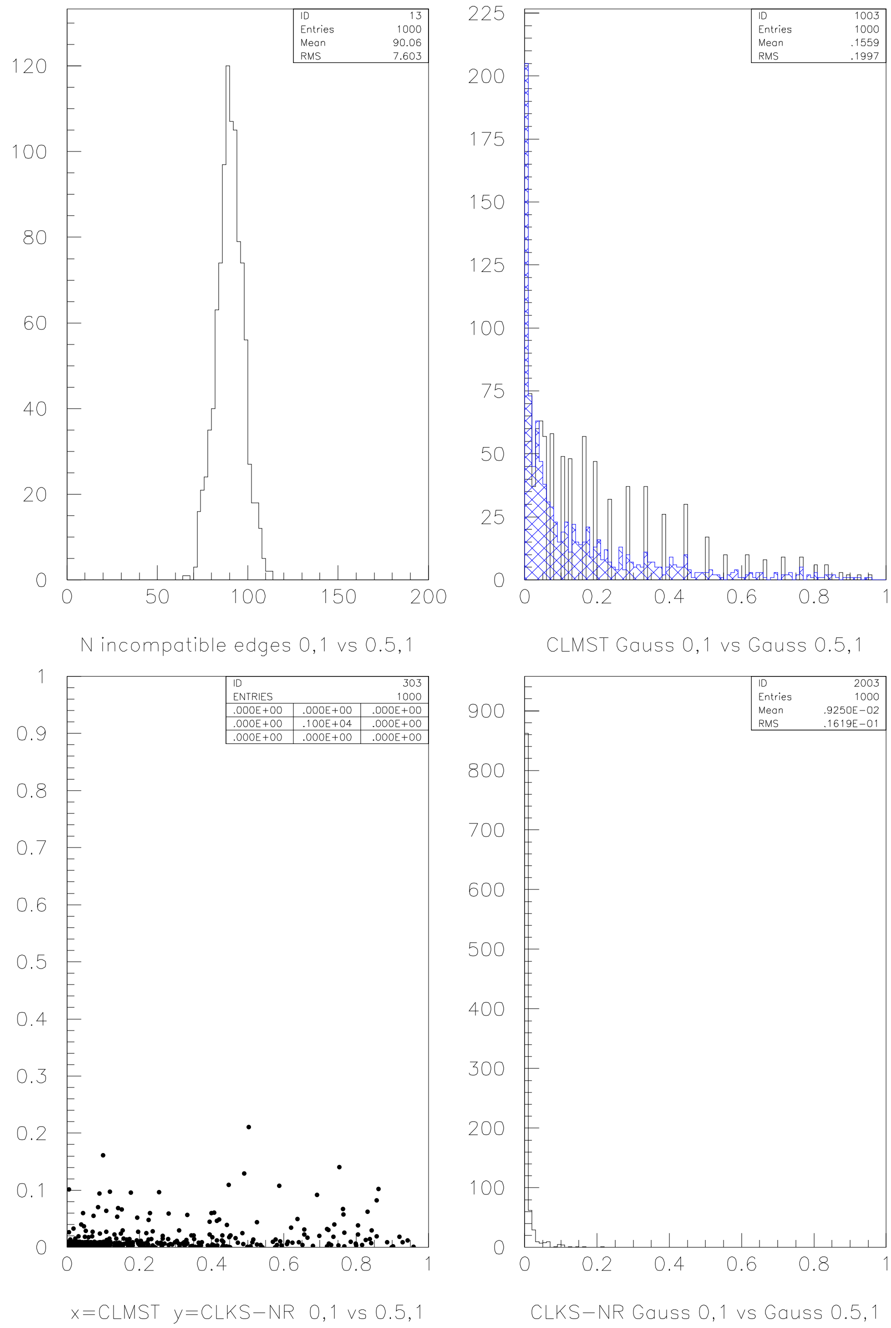


Figure 7: As in Fig. 4, but comparing samples from a 2-dimensional standard normal distribution with samples from a 2-dimensional normal distribution with means 0.5 and standard deviations 1.

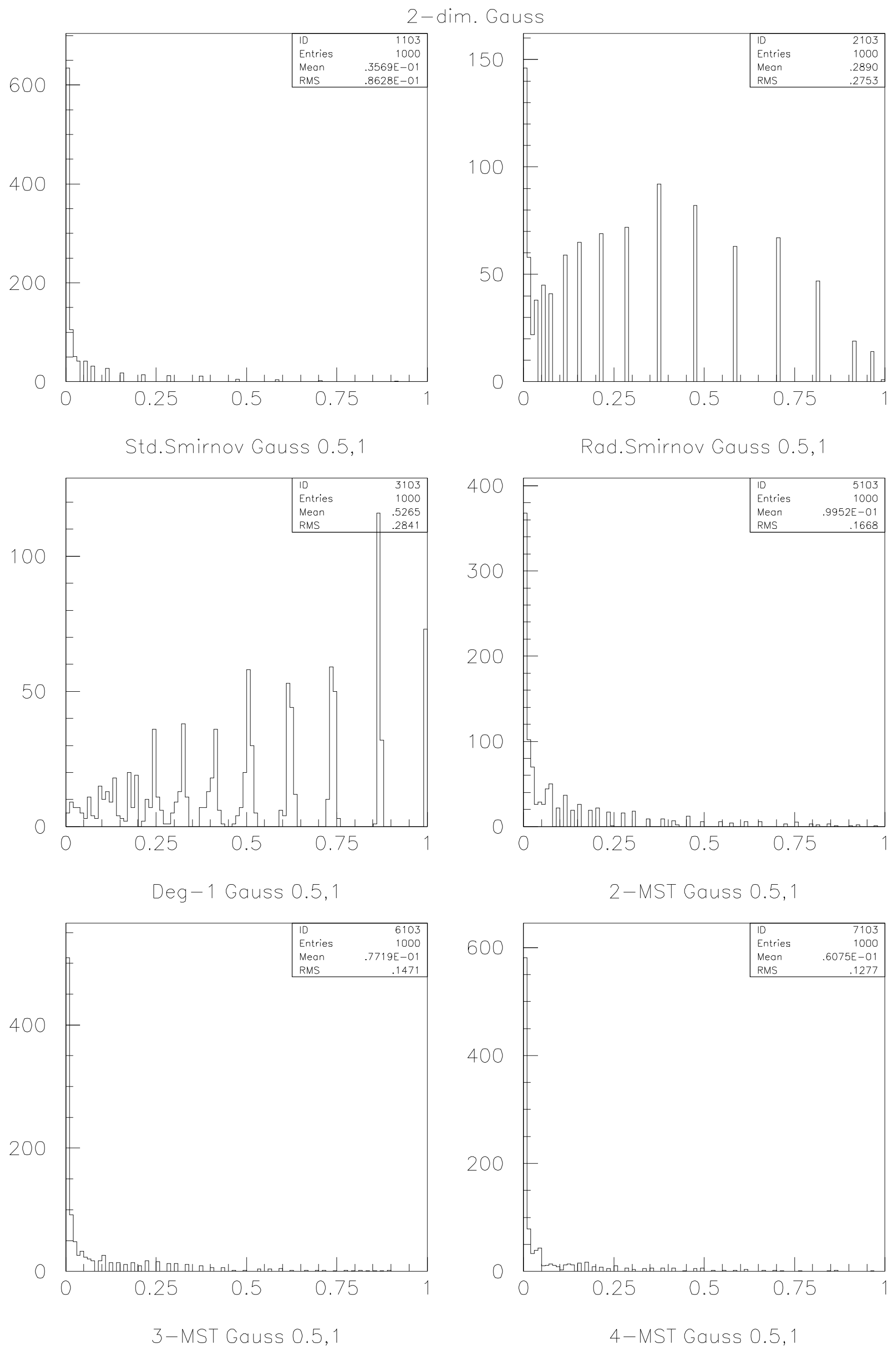


Figure 8: As in Fig. 5, but comparing samples from a 2-dimensional standard normal distribution with samples from a 2-dimensional normal distribution with means 0.5 and standard deviations 1.

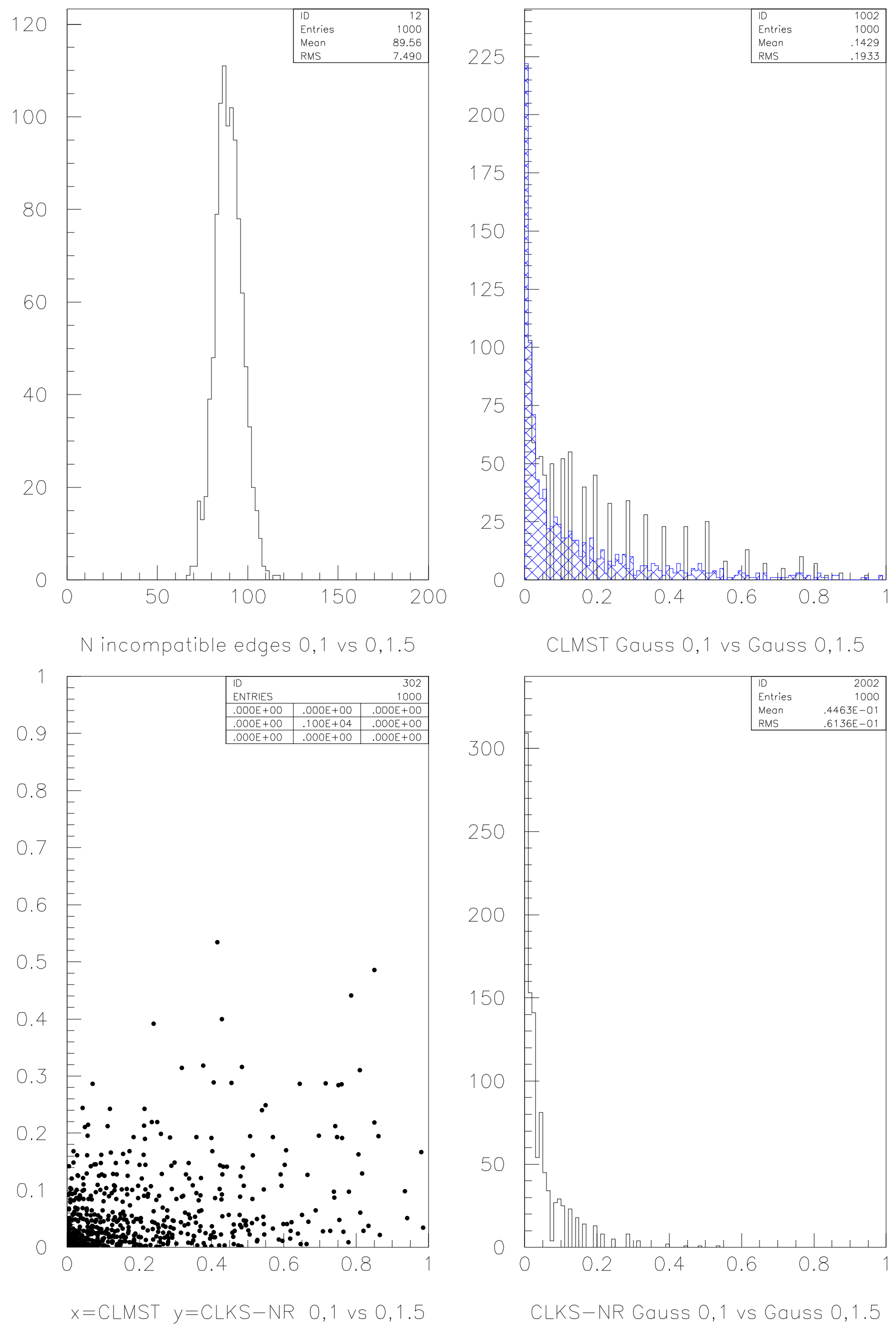


Figure 9: As in Fig. 4, but comparing samples from a 2-dimensional standard normal distribution with samples from a 2-dimensional normal distribution with means 0 and standard deviations 1.5

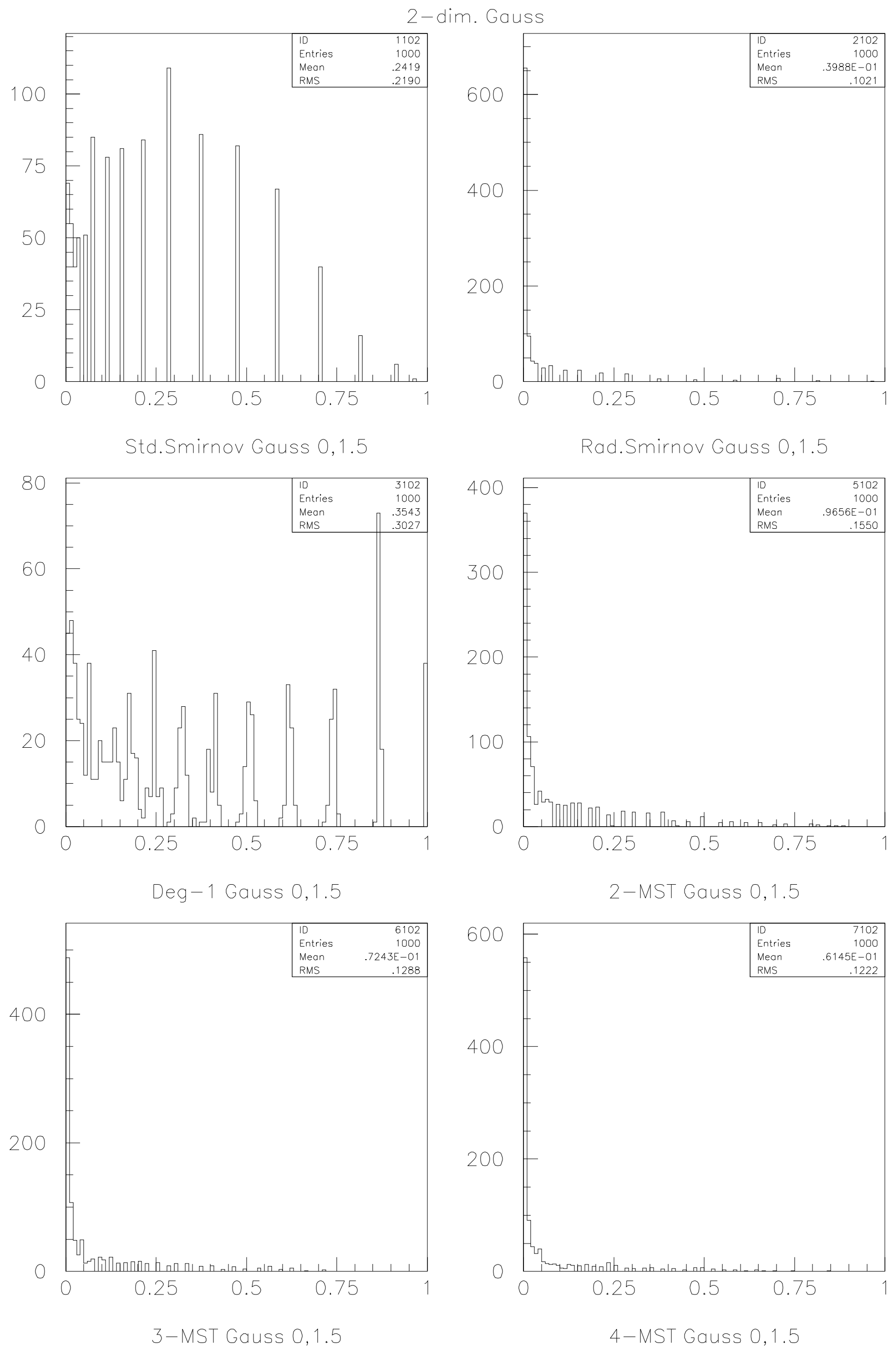


Figure 10: As in Fig. 5, but comparing samples from a 2-dimensional standard normal distribution with samples from a 2-dimensional normal distribution with means 0 and standard deviations 1.5

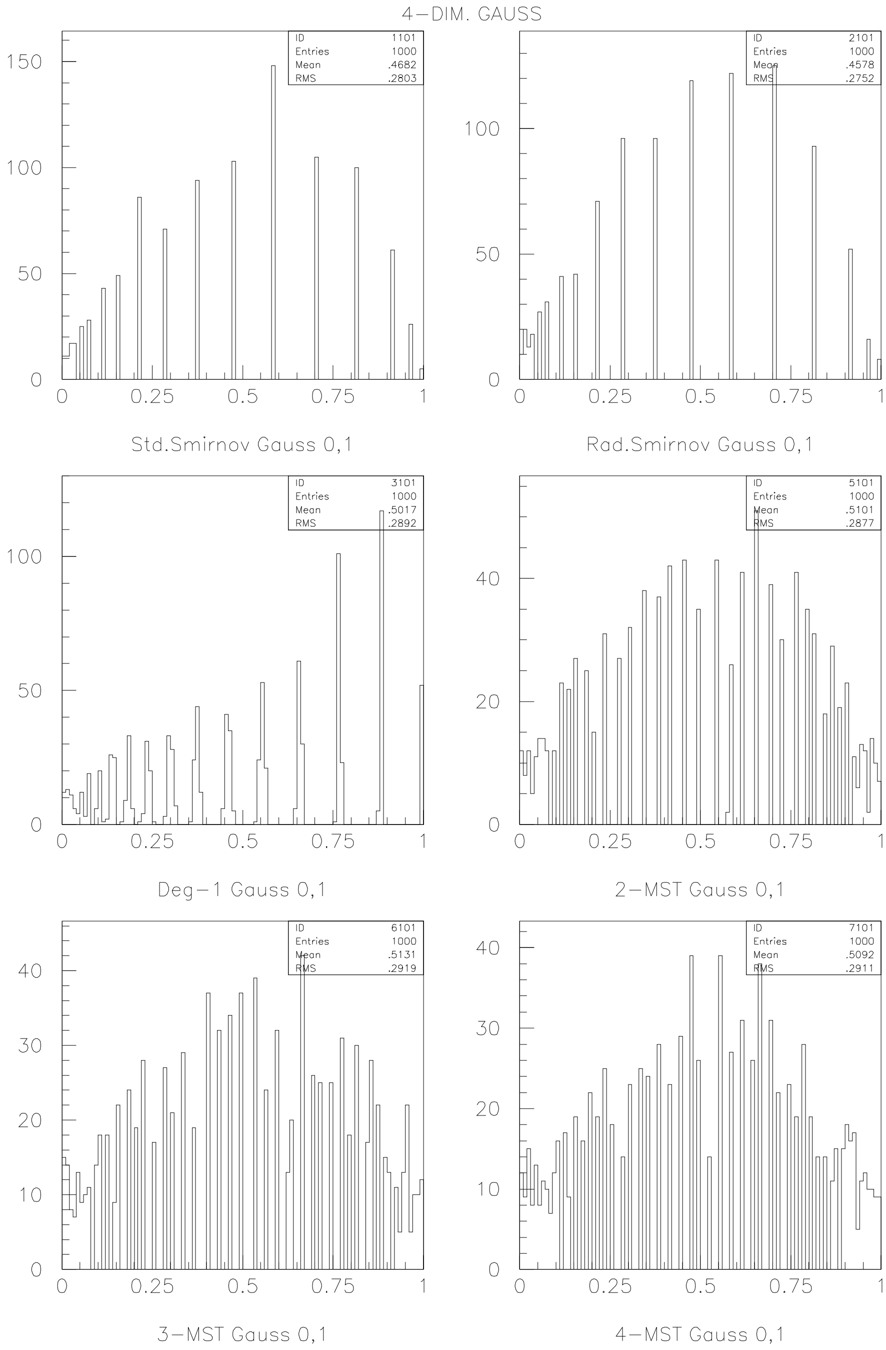


Figure 11: Confidence levels for 2 samples of 100 points, both from a 4-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

Figure 12: Correlations between confidence levels for 2 samples of 100 points, both from a 4-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

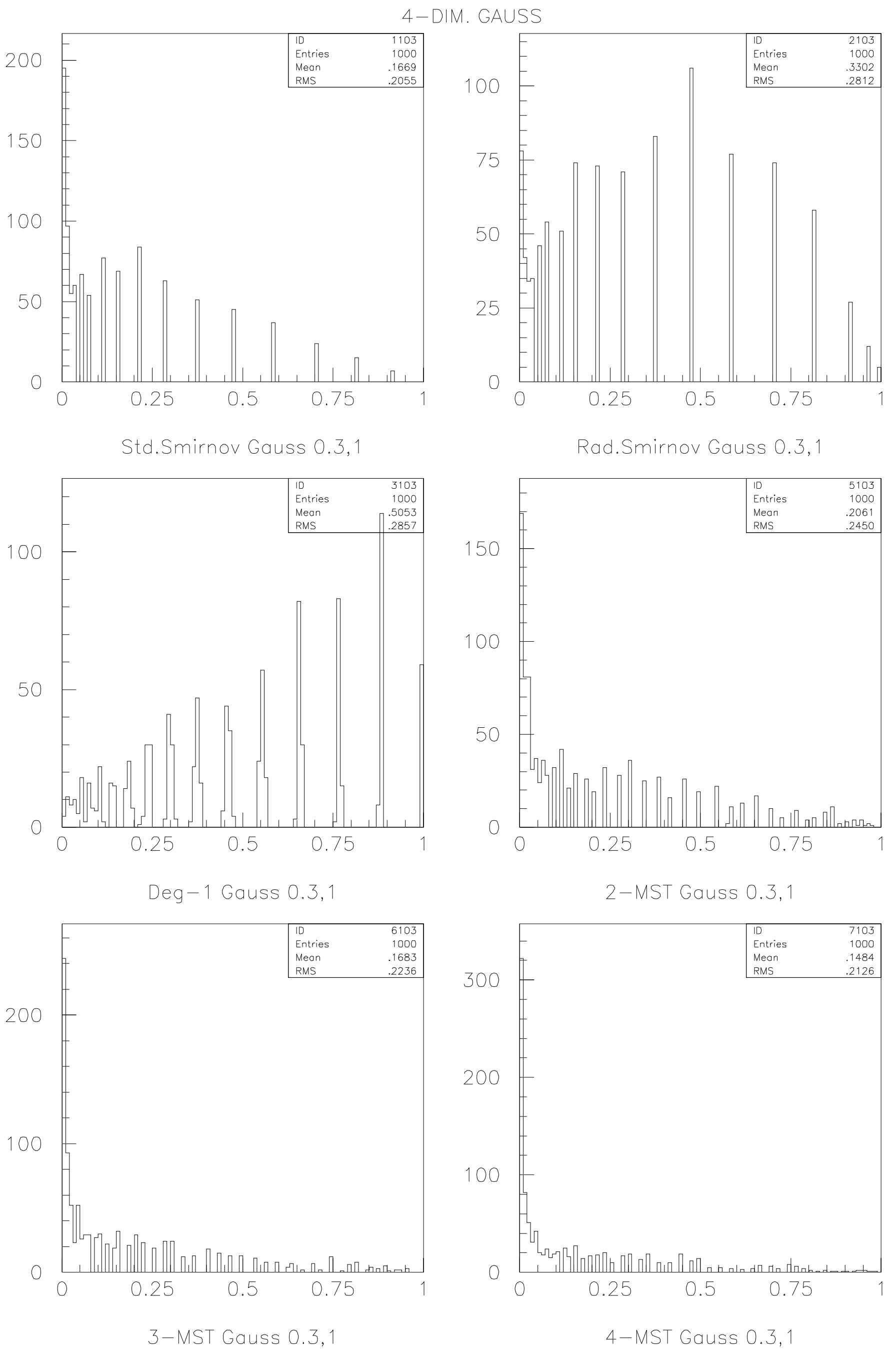


Figure 13: As in Fig. 8, but comparing samples from a 4-dimensional standard normal distribution with samples from a 4-dimensional normal distribution with means 0.3 and standard deviations 1.

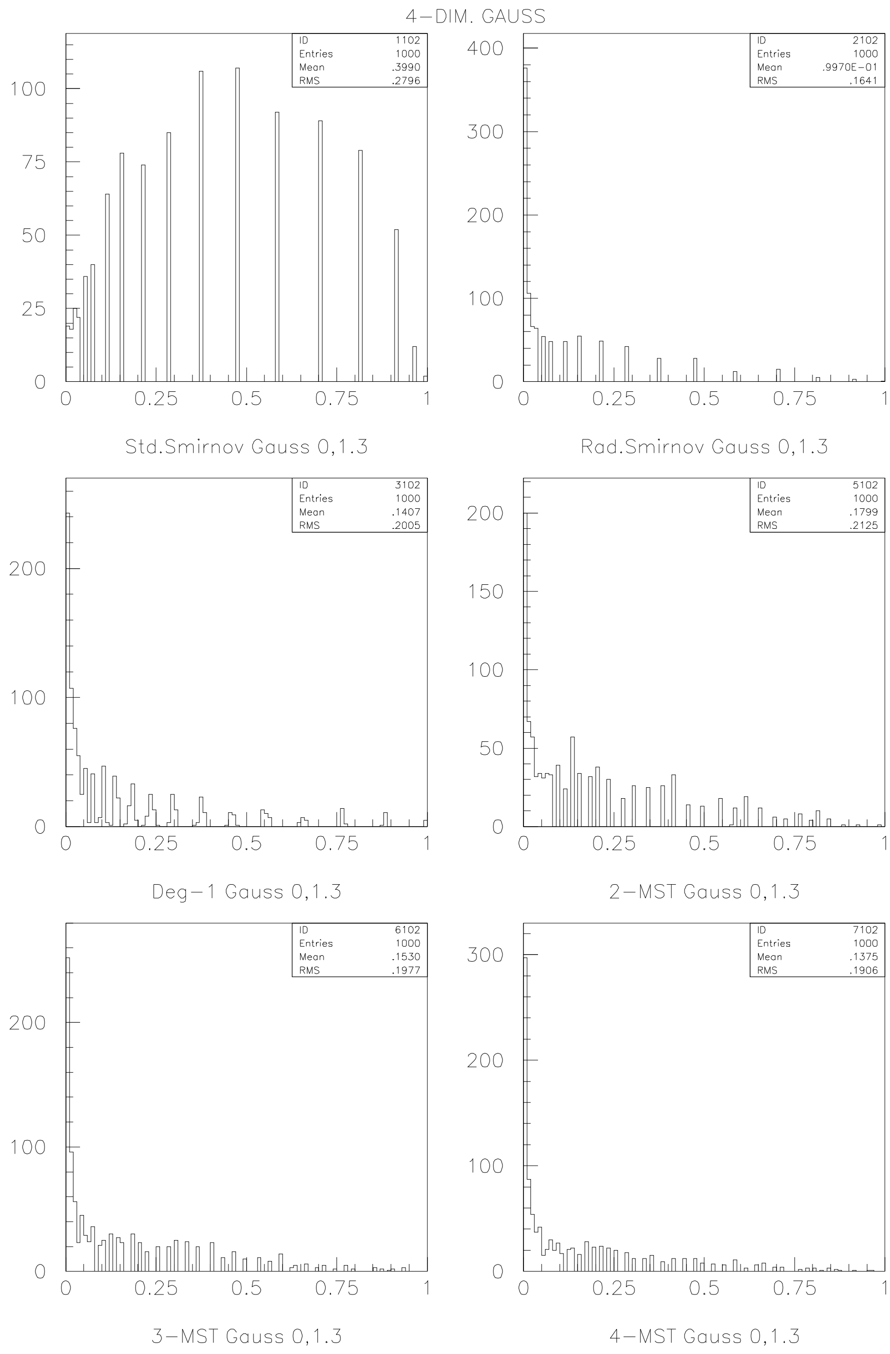


Figure 14: As in Fig. 10, but comparing samples from a 4-dimensional standard normal distribution with samples from a 4-dimensional normal distribution with means 0 and standard deviations 1.3

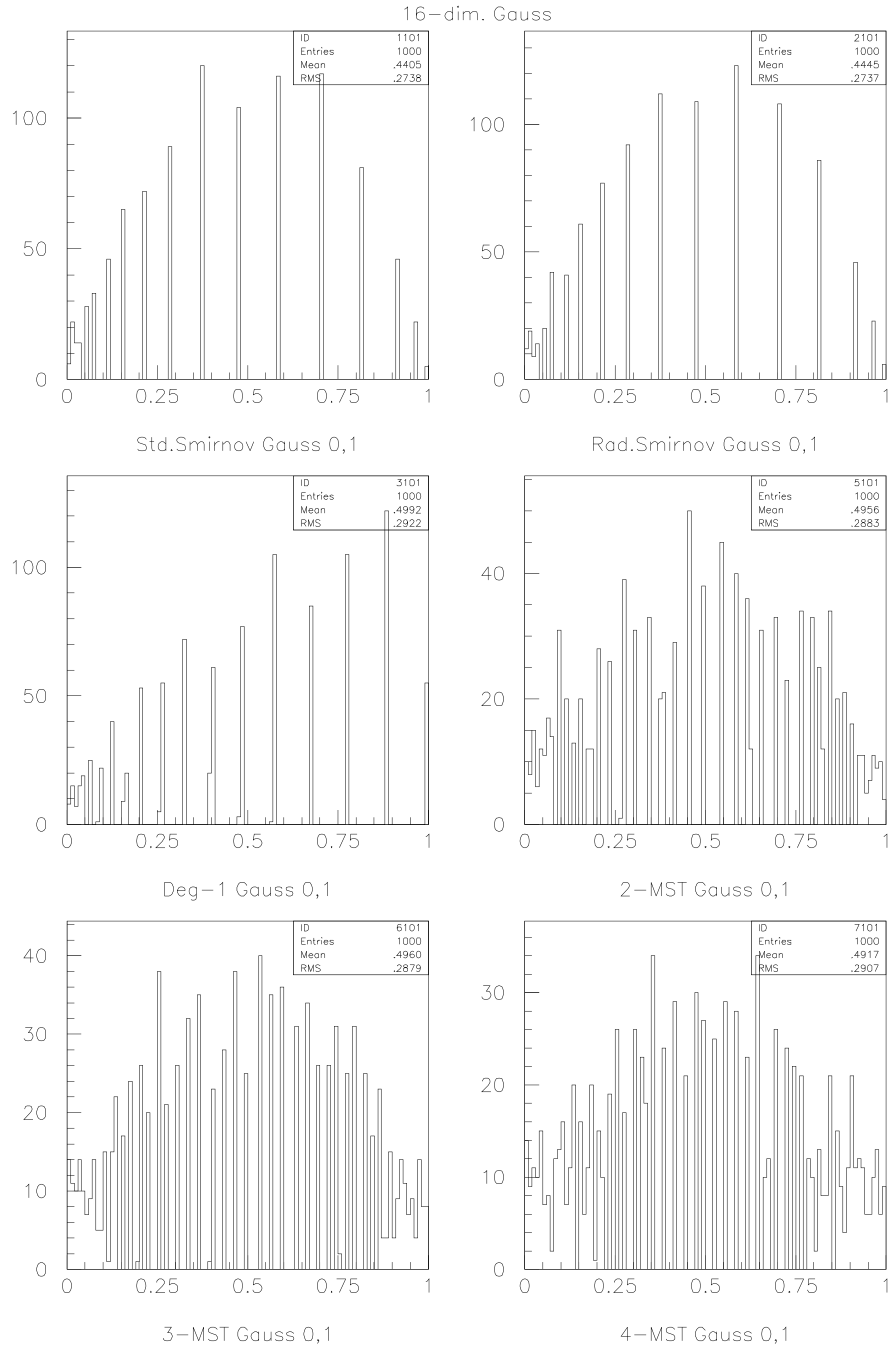


Figure 15: Confidence levels for 2 samples of 100 points, both from a 16-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

Figure 16: Correlations between confidence levels for 2 samples of 100 points, both from a 16-dimensional standard normal distribution, for 1000 pairs of samples calculated using various MST tests:[4] standard Smirnov; radial Smirnov; number of degree-1 nodes; number of incompatible edges in 2 MSTs, 3 MSTs, and 4 MSTs.

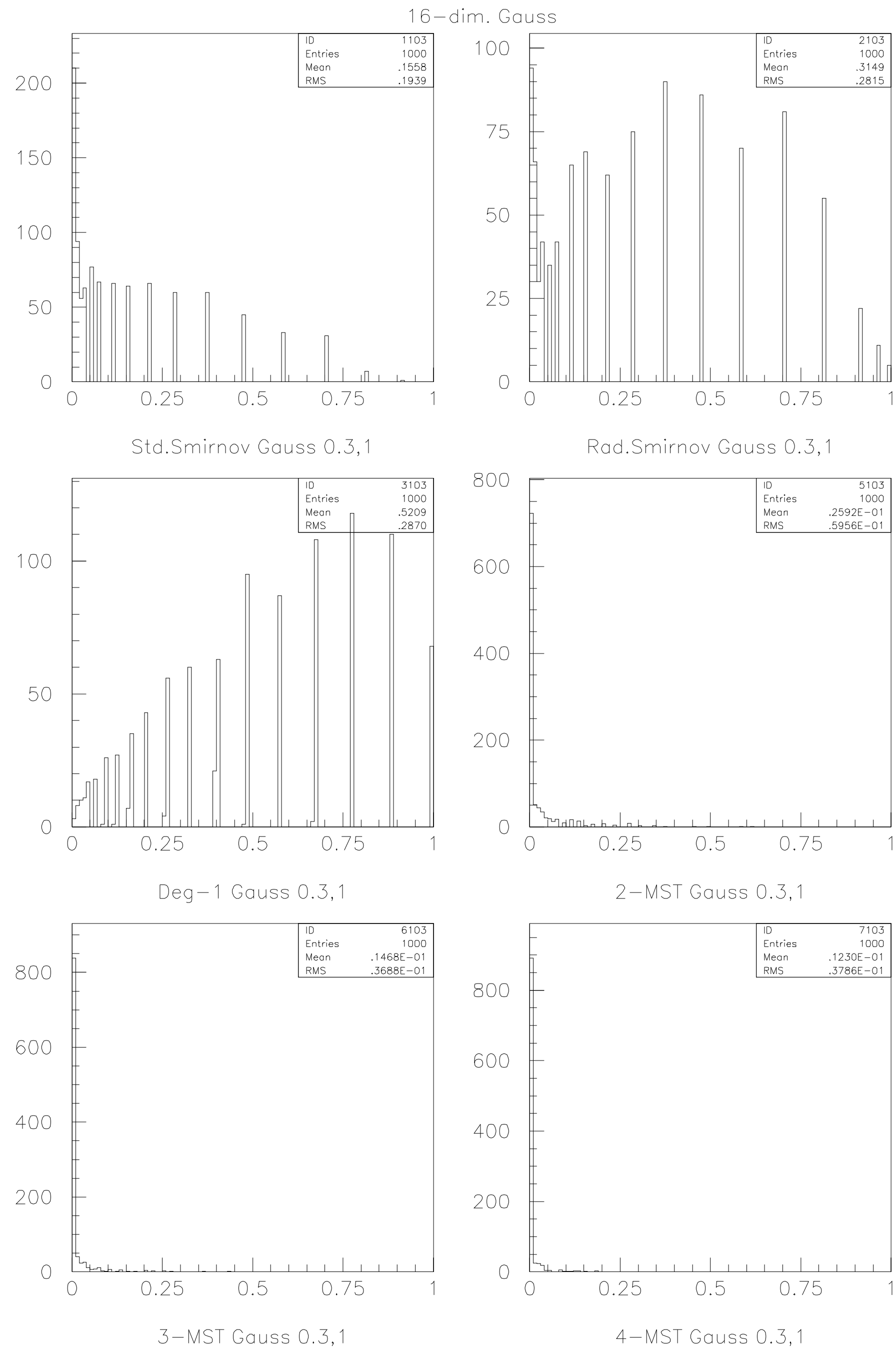


Figure 17: As in Fig. 8, but comparing samples from a 16-dimensional standard normal distribution with samples from a 16-dimensional normal distribution with means 0.3 and standard deviations 1.

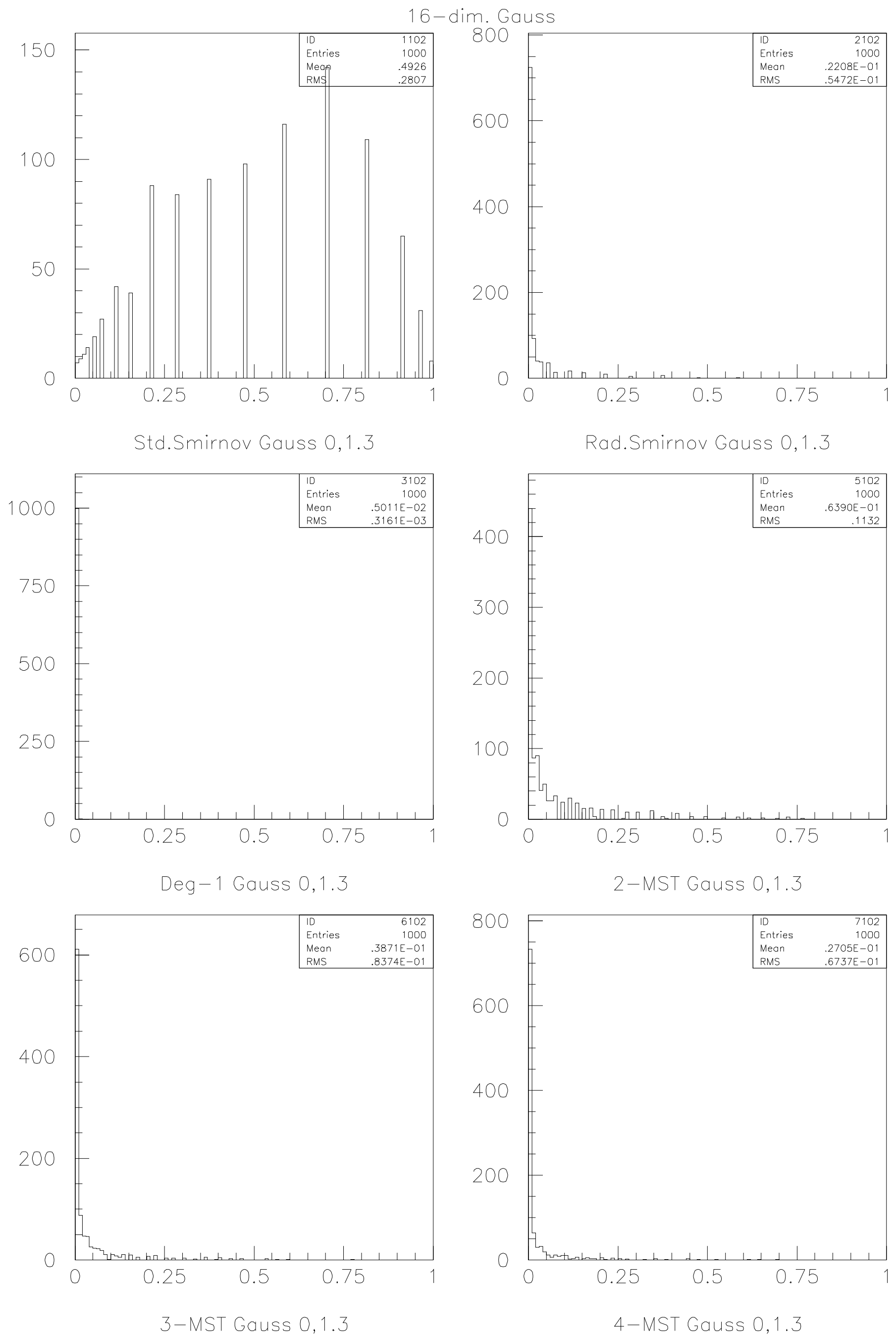


Figure 18: As in Fig. 10, but comparing samples from a 16-dimensional standard normal distribution with samples from a 16-dimensional normal distribution with means 0 and standard deviations 1.3