DIMACS Technical Report 98-45
October 1998

Partition-Based Goodness-of-Fit Tests on the Line and the Circle[1]

by

Carol J. Feltz[2]
Northern Illinois University, USA

Gerald A. Goldin
Rutgers University, New Jersey USA

[1] *In memory of N. H. Kuiper*
[2] *Address for correspondence*: Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA.
E-mail: `feltz@math.niu.edu`

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, Bellcore and NEC Research Institute.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

We propose a way to construct consistent generalizations of the Cramér-von Mises test on the line, and the Watson U_n^2 test on the circle, based on classes of partitions invariant under various group actions. The framework developed extends our earlier work generalizing the Kolmogorov-Smirnov and Kuiper goodness-of-fit tests, and provides a conceptually unifying description. Many goodness-of-fit tests for uniformity on the circle incorporate rotation invariance. Our construction does not require that the distribution for the null hypothesis be uniform, and permits tests invariant under general coordinate transformations. Some distribution-free tests are presented as examples, and their properties explored through numerical simulations.

Key words and phrases: Distribution-free tests; Kolmogorov-Smirnov test; Kuiper's test; Cramér-von Mises test; Watson U_n^2 test.

1. INTRODUCTION

Various adaptations of goodness-of-fit tests from the line $\mathbf{R}^1$ to the circle $\mathbf{S}^1$ have been proposed. In this section we review them, and some other related tests on the circle. This establishes notation for the rest of the paper.

We then introduce a framework for obtaining consistent generalizations of the Cramér-von Mises test on $\mathbf{R}^1$, and the Watson U_n^2 test on $\mathbf{S}^1$, based on classes of partitions invariant under the action of various groups of transformations—including the group of general coordinate transformations (diffeomorphisms). Our approach extends previous work generalizing the Kolmogorov-Smirnov (K-S) and Kuiper goodness-of-fit tests, suggesting additional, powerful tests that can be tailored to the known symmetries of a situation (Feltz and Goldin, 1992). The result is also a conceptual unification of many seemingly disparate or *ad hoc* tests discussed in the literature. Finally, we explore the first few of the natural series of distribution-free tests following from our analysis, using numerical simulation.

Let $f(x)$ be an integrable, nonvanishing probability density function on $\mathbf{R}^1$ or $\mathbf{S}^1$, and $\mathcal{S} = \{x_1, \ldots, x_n\}$ a set of data points. The goal is to test the null hypothesis H_0 that the data $\mathcal{S}$ are drawn randomly from f. On $\mathbf{R}^1$, let $F(x) = \int_{-\infty}^{x} f(u)\, du$ be the cumulative distribution function. Let $F_n^{emp}(x) = n^{-1} \sum \Theta(x - x_i) = \int_{-\infty}^{x} du \{ n^{-1} \sum \delta (u - x_i) \}$ be the sample distribution function, where $\Theta(x) = 1$ for $x \geq 0$ and 0 otherwise, and $\delta(x)$ is the Dirac δ-function. Thus on $\mathbf{R}^1$, $F(-\infty) = F_n^{emp}(-\infty) = 0$, and $F(+\infty) = F_n^{emp}(+\infty) = 1$.

We parameterize $\mathbf{S}^1$ with the variable x ranging from 0 to 1, where the 0 point and the direction around the circle have been chosen arbitrarily; addition on the interval is defined *modulo* 1. On $\mathbf{S}^1$ we have $F(x) = \int_0^x f(u)\, du$, and $F_n^{emp}(x) = \int_0^x du \{ n^{-1} \sum \delta (u - x_i) \}$; with $F(0) = F_n^{emp}(0) = 0$, and $F(1) = F_n^{emp}(1) = 1$.

The K-S statistic on $\mathbf{R}^1$, given by

$$D_n = \sup_{x \in \mathbf{R}^1} | F_n^{emp}(x) - F(x) | , \tag{1.1}$$

was modified by Kuiper (1960) for $\mathbf{S}^1$,

$$V_n = \sup_{x \in [0,1)} \{F_n^{emp}(x) - F(x)\} - \inf_{x \in [0,1)} \{F_n^{emp}(x) - F(x)\} . \tag{1.2}$$

The Cramér-von Mises statistic on $\mathbf{R}^1$ is defined by

$$W_n^2 = n \int_{-\infty}^{+\infty} \{F_n^{emp}(x) - F(x)\}^2 dF(x) ; \tag{1.3}$$

Watson (1961) adapted (1.3) for $\mathbf{S}^1$,

$$U_n^2 = n \int_{x=0}^{1} \{F_n^{emp}(x) - F(x) - \int_{y=0}^{1} [F_n^{emp}(y) - F(y)]\, dF(y)\}^2 dF(x) . \tag{1.4}$$

The latter statistic was regarded by Watson as conceptually similar to a variance. The tests based on (1.1)-(1.4) are all distribution-free and consistent against all alternatives; while (1.2) and (1.4) are also rotation invariant. But, as Watson notes, special tests may be more powerful than such "general purpose tests" against definite classes of alternatives.

For example, Ajne (1968) studied a rotation-invariant test for uniformity on $\mathbf{S}^1$ based on semicircles, using the statistic

$$A_n = n^{-1} \int_0^1 \{N(x) - \frac{1}{2}n\}^2 dx\,, \tag{1.5}$$

where $N(x)$ is the number of data points falling in the semicircular interval $[\,x, x+\frac{1}{2})$. The distribution of N under the assumption of uniformity is a purely combinatorial problem, addressed by Ajne using a random-walk argument. Watson (1967a) derived the asymptotic null distribution of A_n. While Ajne's test is obviously not consistent against all alternative distributions, it is quite powerful against a certain class of alternatives: those that are uniform with density $2p$ on some semicircle, and uniform with density $2q$ (where $q = 1 - p$) on the complementary semicircle. A general form for the rotation-invariant test for uniformity on $\mathbf{S}^1$ that is most powerful against such alternatives was investigated by Ajne. He found A_n to be optimal in testing against a set of local alternatives (p and q near $1/2$). Ajne also considered the rotation-invariant statistic

$$N = \max_{x \in [0,1)} N(x)\,, \tag{1.6}$$

which can be understood as an adaptation to $\mathbf{S}^1$ of the K-S statistic, based on semicircles. Although a test using (1.6) is also not consistent against all alternatives, it is optimal against distant alternatives in the above-described class.

Beran (1968, 1969) extended the ideas relating to (1.4) and (1.5) to the case of a compact homogeneous space, and developed the asymptotic theory of such tests. He observed that (1.4) and (1.5), as well as a statistic of Lord Rayleigh,

$$R_n^2 = n^{-1} \left[\left(\sum_{j=1}^n \cos 2\pi x_j \right)^2 + \left(\sum_{j=1}^n \sin 2\pi x_j \right)^2 \right], \tag{1.7}$$

and another statistic of Watson (1967b),

$$S_n{}^2 = n^{-1} \int_0^1 \left[f_n(x) - 1 \right]^2 \, dx \tag{1.8}$$

(where f_n is a consistent estimator of the true density from which the data are drawn), are all special cases of the statistic,

$$T_n = n^{-1} \int_0^1 \left[\sum_{j=1}^n f(x + x_j) - n \right]^2 dx \tag{1.9}$$

(where f is a square-integrable probability density on the circle). Beran described the asymptotic distribution of T_n under fixed alternatives to uniformity on $[0,1]$, and under sequences of local alternatives. He determined the alternative distributions against which T_n (for f specified) provides a consistent test for uniformity.

Rothman (1972) developed a more general form of (1.5) using arcs of arbitrary length, having the property of consistency. In its most general form, Rothman's statistic is given by

$$A_n^{H(t)} = \int_0^1 \int_0^1 [N(t,x) - nt]^2 \, dx \, dH(t) \tag{1.10}$$

where $H(t)$ is a distribution function, and $N(t,x)$ is the number of observations in the arc $(x, x+t]$. When $H(t) = t$, the uniform distribution, then (1.10) is equivalent to (1.4). To demonstrate this, Rothman expands $N(t,x) - nt$ as a Fourier series, but there is a much more direct proof which we offer in Eqs. (3.2)-(3.5).

The fact that Watson's statistic is a special case of Rothman's is noteworthy, since (1.4) can also be written

$$U_n^2 = \inf_{x_0 \in \mathbf{S}^1} n \int_0^1 \{F_n^{emp}(x;x_0) - F(x;x_0)\}^2 dF(x) \tag{1.11}$$

where $F_n^{emp}(x;x_0)$ is the empirical distribution function calculated with x_0 as the starting point, and $F(x;x_0)$ is the corresponding theoretical distribution function. That is, Watson was in effect choosing the starting point on the circle so as to minimize the value of the Cramér-von Mises statistic (thinking of this as minimizing a variance); while Rothman obtained (1.10) by averaging over a mixture of Ajne-type statistics for various arc lengths. It is the latter perspective that we generalize in this paper.

Watson (1974) further discussed supremum-type tests on $\mathbf{S}^1$ for distant alternatives to uniformity. In this work, the question of the optimality of (1.2) remains open. Subsequently Cressie (1977) considered the circular scan statistic, defined using Rothman's notation to be

$$N_t = \sup_{0 \le x \le 1-t} N(t,x). \tag{1.12}$$

He showed that if the data come from a uniform distribution, then N_t (properly normalized) converges in distribution to a Brownian motion process. In this the length t of the arc is fixed. However, Cressie also demonstrated that with proper normalization, taking the supremum over t of N_t leads to Kuiper's statistic.

Other approaches to testing uniformity on the circle include the work of Rao (1976), Kokic (1987), and Huffer (1988), who developed tests based on the arc length of the spacings between data points, and the work of Hermans and Rasson (1985), who proposed a test using the work of Gine (1975) based on Sobolev norms. Ahmad (1993) modified Watson's statistic (1.4), to obtain asymptotically normal null distributions.

Our earlier work (Feltz and Goldin, 1992) included a series of consistent generalizations $D_n^{(k)}$ of the K-S test on $\mathbf{R}^1$, which we called k-point tests ($k = 1, 2, 3, \ldots$).

The case $k = 1$ is the K-S test (1.1), and the case $k = 2$ turns out to be exactly the analogue on $\mathbf{R}^1$ of Kuiper's test on $\mathbf{S}^1$, with the same critical values (though we did not then note this fact). We estimated critical values for $D_n^{(2)}$, $D_n^{(3)}$, and $D_n^{(4)}$ with Monte-Carlo techniques, and presented the results of a power study.

The present paper extends this work to obtain systematic generalizations of the Cramér-von Mises test on $\mathbf{R}^1$, and Watson's U_n^2 test on $\mathbf{S}^1$, based on classes of partitions invariant under various group actions. Thus we have, in particular, two series of consistent, distribution-free, partition-based tests on the line (k-point tests, for $k = 1, 2, 3, \ldots$), and two series on the circle (k-interval tests, for $k = 1, 2, 3, \ldots$). Tests based on the sup-norm generalize the K-S test on $\mathbf{R}^1$ and Kuiper's test on $\mathbf{S}^1$, while tests based on the L^2-norm generalize the Cramér-von Mises test on $\mathbf{R}^1$ and Watson's test on $\mathbf{S}^1$. The distribution on $\mathbf{S}^1$ being tested for goodness of fit is entirely general, so using the probability integral transformation to change the density to a uniform distribution (and subsequently to move the data points) is not necessary.

In the next section we discuss partition-based tests and group invariance for $\mathbf{S}^1$, developing as examples the category of k-interval supremum-based tests. In Sec. 3, we show how to extend partition-based tests and group invariance to obtain analogous generalizations of the Cramér-von Mises and Watson tests. Sec. 4 reports computational formulas and Sec. 5 reports critical values for some of the generalizations described in Sec. 3. Finally, in Sec. 6 we investigate and discuss the power of some of our tests against various alternatives, identifying certain alternatives where the new tests are considerably more powerful.

2. PARTITION-BASED TESTS AND GROUP INVARIANCE FOR $\mathbf{S}^1$

In this section, we develop the framework and notation for partition-based tests for $\mathbf{S}^1$, using Kuiper's V_n statistic as an example. To express the Kuiper test in terms of partitions, suppose we have partitioned $\mathbf{S}^1$ into two Borel subsets, $B_1^{a,b}$ and $B_2^{a,b}$, where $B_1^{a,b}$ is the half-open interval of arc from a to b, and $B_2^{a,b}$ is the complementary arc from b to a; so that $B_1^{a,b} \bigcup B_2^{a,b} = \mathbf{S}^1$ and $B_1^{a,b} \bigcap B_2^{a,b} = \emptyset$. Call such a partition $\pi^{a,b}$. Let Π be the family of all such measurable partitions of $\mathbf{S}^1$, so $\Pi = \{\pi^{a,b}$ for $a, b \in \mathbf{S}^1\}$.

Then, redefining the K-S metric in terms of partitions,

$$\kappa_{n,f}^{\Pi}(\mathcal{S}) = sup_{\{\pi^{a,b} \in \Pi\}} \left| \frac{n_{\pi^{a,b}}(\mathcal{S})}{n} - F_{\pi^{a,b}} \right| , \tag{2.1}$$

where $F_{\pi^{a,b}} = \int_{B_1{}^{a,b}} f(x)dx$, $n = \#(\mathcal{S})$ and $n_{\pi^{a,b}}(\mathcal{S}) = \#(\mathcal{S} \bigcap B_1{}^{a,b})$. Notice that $\kappa_{n,f}^{\Pi}(\mathcal{S})$ depends on the family of partitions Π as well as on the sample $\mathcal{S}$ and the density f in its calculation. This redefinition of the K-S metric to accommodate it to $\mathbf{S}^1$ is equivalent to the Kuiper statistic when $f(x)$ is the uniform distribution, as pointed out by Barr and Shudde (1973). The Kuiper statistic is also equivalent to a

2-point K-S test (interval test) on the line, as constructed in Feltz and Goldin (1992). We can now adapt our earlier methods for $\mathbf{R}^1$ straightforwardly to $\mathbf{S}^1$.

In generalizing the Kuiper statistic on $\mathbf{S}^1$, we focus attention on the situation where there exists a group G of coordinate transformations of $\mathbf{S}^1$ under which Π is invariant. That is, for all $T \in G$, if $(B_1, B_2) \in \Pi$, then $(TB_1, TB_2) \in \Pi$. Natural candidates for such invariance groups in $\mathbf{S}^1$ include groups of rotations, Möbius transformations, and diffeomorphisms. If the invariance group is smaller than the full group of diffeomorphisms, the distribution of the statistic $\kappa^{\Pi}_{n,f}(\mathcal{S})$ is in general invariant not over all densities, but within families of densities (orbits). Letting f_0 be any particular density, such a family is specified as the G-orbit $\Delta = \{Tf_0 : T \in G\}$. For $G = \mathit{Diff}(\mathbf{S}^1)$, the whole space of smooth densities on $\mathbf{S}^1$ is a single orbit under G, so that $\kappa^{\Pi}_{n,f}(\mathcal{S})$ is distribution-free. When G is some other group acting on $\mathbf{S}^1$, there can be many distinct orbits.

Evidently $\Pi = \{\pi^{a,b}\}$ is invariant under $\mathit{Diff}(\mathbf{S}^1)$. A family of distribution-free, K-S type tests on $\mathbf{S}^1$ is obtained from the family of k-interval statistics, with Kuiper's test corresponding to $k = 1$. Let $(a_1, a_2, \ldots, a_{2k})$ be an ordered set of $2k$ points in $\mathbf{S}^1$; let $\pi^{a_1,a_2,\ldots,a_{2k}} = (B_1^{a_1,a_2,\ldots,a_{2k}}, B_2^{a_1,a_2,\ldots,a_{2k}})$, where $B_1^{a_1,a_2,\ldots,a_{2k}} = (a_1, a_2] \bigcup (a_3, a_4] \bigcup \ldots \bigcup (a_{2k-1}, a_{2k}]$; and let $\Pi = \{\pi^{a_1,a_2,\ldots,a_{2k}}\}$. The form of the corresponding statistic looks the same as in (2.1), but the supremum is taken over a larger set of partitions, creating a new test for each distinct k. For k fixed, each family Π of k-interval partitions is a Vapnik-Chervonenkis class (Vapnik and Chervonenkis, 1971). This means that the relative frequency of data in each subset in the class converges to the probability of that subset uniformly over the class—heuristically, there are not so many partitions that the supremum is identically 1. However, because the supremum is taken over all possible endpoints, there are enough partitions to make each of the tests based on these families of partitions consistent as well as distribution-free.

Examples of tests that are not distribution-free, and also not consistent against all alternatives, are those making use of statistics based on partitions of fixed interval width, or having a fixed range of interval widths. One such example is Ajne's A_n statistic using semicircular intervals of $\mathbf{S}^1$. To generalize this, let L be a fixed positive real number less than 1/2 (the fixed width). Let $\pi^{a,a+L} = ({B_1}^{a,a+L}, {B_2}^{a,a+L})$; i.e., take partitions of $\mathbf{S}^1$ by half-open arcs from a to $a + L$. Let $\Pi^L = \{\pi^{a,a+L} : a \in \mathbf{S}^1\}$. This family of partitions is also a Vapnik-Chervonenkis class. But the test statistic based on Π^L is invariant only under rotations of $\mathbf{S}^1$ (or, correspondingly, under translations of $\mathbf{R}^1$, for any fixed L), not under general diffeomorphisms since dilations can change L. While families of partitions of fixed width are not large enough to give tests that are consistent, the resulting tests will (like Ajne's) be optimal against particular alternatives.

3. GENERALIZATION OF THE CRAMÉR-VON MISES STATISTIC FOR $\mathbf{S}^1$ AND $\mathbf{R}^1$

As discussed in the introduction, Watson created a distribution-free analogue of the Cramér-von Mises test on $\mathbf{S}^1$. Squaring out the integrand in equation (1.4), we have

$$U_n^2 = n\int_0^1 \{F^{emp}(x) - F(x)\}^2 dF(x) - n\{\int_0^1 [F^{emp}(x) - F(x)]\, dF(x)\}^2. \qquad (3.1)$$

Next we write U_n^2 in terms of partitions, using the notation of the previous section:

$$\begin{aligned}
2\,U_n^2 &= n\int_0^1\int_0^1 \{F_n^{emp}(b) - F(b) - [F_n^{emp}(a) - F(a)]\}^2 dF(b)dF(a) && (3.2)\\
&= n\int_0^1\int_0^1 \{F_n^{emp}(b) - F_n^{emp}(a) - [F(b) - F(a)]\}^2 dF(b)dF(a). && (3.3)\\
&= n\int_0^1\int_0^1 \left(\frac{n_{\pi^{a,b}}(\mathcal{S})}{n} - \int_{B_1{}^{a,b}} f(x)dx\right)^2 f(b)db f(a)da && (3.4)\\
&= n\int_0^1\int_a^1 \left(\frac{n_{\pi^{a,b}}(\mathcal{S})}{n} - \int_{B_1{}^{a,b}} f(x)dx\right)^2 f(b)db f(a)da \\
&\quad + n\int_0^1\int_0^a \left(1 - \frac{n_{\pi^{a,b}}(\mathcal{S})}{n} - (1 - \int_{B_1{}^{a,b}} f(x)dx)\right)^2 f(b)db f(a)da\,. && (3.5)
\end{aligned}$$

Letting $f(x)$ be the uniform density in the above, we reproduce Rothman's result mentioned in the introduction without a Fourier expansion.

We also make the interesting observation that Watson's U_n^2 statistic is equivalent to the 2-point generalization of the Cramér-von Mises statistic for the line, which can be written

$$CvM_2 = n\int_{-\infty}^{\infty}\int_a^{\infty} \left(\frac{n_{\pi^{a,b}}(\mathcal{S})}{n} - \int_a^b f(x)dx\right)^2 f(b)db f(a)da\,. \qquad (3.6)$$

In a similar way we can create a 3-point generalization of the Cramér-von Mises statistic on the line:

$$CvM_3 = n\int_{-\infty}^{\infty}\int_a^{\infty}\int_b^{\infty} \left(\frac{n_{\pi^{a,b,c}}(\mathcal{S})}{n} - (\int_{-\infty}^{a} f(x)dx + \int_b^c f(x)dx)\right)^2 f(c)dc f(b)db f(a)da\,, \qquad (3.7)$$

where the partitions are of the form $\pi^{a,b,c} = (B_1^{a,b,c}, B_2^{a,b,c})$, with $B_1^{a,b,c} = (-\infty, a] \bigcup (b, c]$ and $B_2^{a,b,c} = (a, b] \bigcup (c, \infty)$. We also introduce a 4-point generalization of the Cramér-von Mises test for the line:

$$\begin{aligned}
CvM_4 &= n\int_{-\infty}^{\infty}\int_a^{\infty}\int_b^{\infty}\int_c^{\infty} \left(\frac{n_{\pi^{a,b,c,d}}(\mathcal{S})}{n} - (\int_a^b f(x)dx + \int_c^d f(x)dx)\right)^2 \\
&\qquad \cdot dF(d)dF(c)dF(b)dF(a) \qquad (3.8)
\end{aligned}$$

where the partitions now are based on subsets of the form $B_1^{a,b,c,d} = (a,b] \bigcup (c,d]$. The additional densities used in the integration are necessary for these statistics to be distribution-free. The 4-point generalization of the Cramér-von Mises test can also be used directly as a 2-interval test on the circle. The computational forms of these statistics are given in the next section.

On $\mathbf{S}^1$, Watson noted that U_n^2 is based on choosing the 0 point that minimizes the squared difference between the empirical and theoretical densities. One can create another analogue of Watson's U_n^2 on the circle by choosing the point x_0 which *maximizes* the squared difference between the empirical and theoretical:

$$S_n^2 = \max_{x_0} \; n \int_0^1 \{F_n^{emp}(x; x_o) - F(x; x_o)\}^2 dF(x)\,. \tag{3.9}$$

A test based on this statistic is also distribution-free.

4. COMPUTATIONAL FORMULAS FOR 3- AND 4-POINT CRAMÉR-VON MISES TESTS

The computational formulas for the above 3- and 4-point tests are somewhat complicated. Let CvM_1 be the computational formula for the Cramér-von Mises test,

$$CvM_1 = n \int_{-\infty}^{\infty} \{F_n^{emp}(x) - F(x)\}^2 dF(x) = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_{(i)}) - \frac{2i-1}{2n} \right)^2 . \tag{4.1}$$

In the notation of Watson (1961), let $v_i = F(x_{(i)})$. Then the computational formula for the "average deviation" of the empirical from the theoretical distribution is

$$\int_{-\infty}^{\infty} [F_n^{emp}(x) - F(x)] dF(x) = \frac{1}{2} - \frac{1}{n} \sum_{i=1}^{n} F(x_{(i)}) = \frac{1}{2} - \bar{v}\,. \tag{4.2}$$

The computational formula for the 2-point Cramér-von Mises statistic is the same as that obtained by Watson for U_n^2,

$$CvM_2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[v_i - \frac{2i-1}{2n} - \left(\bar{v} - \frac{1}{2} \right) \right]^2 . \tag{4.3}$$

With some work, the computational formula for the 3-point Cramér-von Mises test can be shown to be:

$$CvM_3 = \frac{1}{2} CvM_1 - n\left(\frac{1}{2} - \bar{v}\right)^2 + 2\xi\,, \tag{4.4}$$

where

$$\xi = \frac{n}{15} - \frac{1}{3}n\bar{v} + \frac{1}{2}\sum_{i=1}^{n}(v_i)^2 - \frac{2}{3n}\sum_{i=1}^{n}(i+1)(v_i)^3$$
$$+\frac{1}{6}\sum_{i=1}^{n}(v_i)^4 + \frac{1}{n}\sum_{i=1}^{n}v_i\sum_{j=1}^{n}(v_j)^2\sigma(j-i)\,, \tag{4.5}$$

and where $\sigma(j-i) = +1$ when $j \geq i$, $\sigma(j-i) = -1$ when $j < i$. Finally the computational formula for the 4-point Cramér-von Mises test is:

$$CvM_4 = \frac{1}{6}CvM_1 - \frac{n}{2}(\frac{1}{2} - \bar{v})^2 + 2\xi$$
$$-4(\frac{1}{2} - \bar{v})[\frac{n}{12} - \frac{1}{3}\sum_{i=1}^{n}(v_i)^3] + n[\frac{1}{3} - \frac{1}{n}\sum_{i=1}^{n}(v_i)^2]^2. \tag{4.6}$$

Equations (4.4)-(4.6) were derived by performing the integrals using "brute force" calculations, simplifying where possible with formulas such as those in Eqns. (4.2) and (4.3). It is interesting that the level of computational complexity of the k-point Cramér-von Mises statistic goes up only with every pair of additional points. The 1-point Cramér-von Mises statistic and the 2-point Cramér-von Mises statistic (Watson's U_n^2) have the same level of computational complexity—they require going through the data once. Similarly, the 3-point and the 4-point Cramér-von Mises statistics have the same level of computational complexity, requiring two iterations through the data. In contrast, the generalizations of the K-S statistic go up a level of computational complexity with each new point—the 1-point K-S test requires cycling through the data once, the 2-point K-S test requires two iterations, and so on.

5. CRITICAL VALUES FOR THE 3- AND 4-POINT CRAMÉR-VON MISES TEST

Here we present critical values for the 3- and 4-point generalizations of the Cramér-von Mises test. These statistics are distribution-free, and so can be used to test the goodness-of-fit of data with any distribution on the line. Critical values for the 4-point generalization of the Cramér-von Mises test can be used to test the goodness-of-fit of data with any distribution on the circle.

Monte Carlo techniques were used to estimate critical values for $\alpha = .10$, $\alpha = .05$, $\alpha = .025$, and $\alpha = .01$, for the sample sizes $2-20$, 25, 30, 35, 40, 45 and 50. Pseudorandom data were generated from a uniform $(0,1)$ distribution, and each critical value was computed for each of 30 separate runs of $20,000$ experiments, so that the variability of the results could be monitored. The average critical values found over the 30 runs for the 3-point Cramér-von Mises test are summarized in Table 1.

The average critical values found over the 30 runs for the 4-point Cramér-von Mises test are summarized in Table 2.

INSERT TABLE 1 AND TABLE 2 APPROXIMATELY HERE

As a cross-check, we used the same Monte-Carlo method to estimate critical values for Kuiper's test and Watson's U_n^2 test. Our critical values for these two statistics were in agreement with those of Stephens (1965) and Stephens (1963).

6. POWER INVESTIGATIONS

To investigate the power of the partition-based Cramér-von Mises tests and the partition-based K-S tests we used as alternatives to uniformity the distributions considered by Stephens (1969). Stephens investigated the power of Ajne's test, Watson's U_n^2 test and Kuiper's test with respect to various uniform distributions on semi-circles:

$$H_A(r,s):\ f(x) = \begin{cases} r & 0 \le x < .5 \\ s & .5 \le x < 1.0 \end{cases} \tag{6.1}$$

where $r/s = 1.2$, 1.5 and 2.0. In addition Stephens looked at the power of the above tests under three clustering alternatives, H_B, H_C, and H_D, defined as follows:

$$H_B(k):\ f(x) = \begin{cases} k(2x)^{k-1} & 0 \le x < .5 \\ k(2(1-x))^{k-1} & .5 \le x < 1.0\,. \end{cases} \tag{6.2}$$

Notice that this density has a peak at $x = .5$ when $k > 1$; for high values of k, points should be clustered around $x = .5$.

$$H_C(k):\ f(x) = \begin{cases} k(4x)^{k-1} & 0 \le x < .25 \\ k(4(.5-x))^{k-1} & .25 \le x < .5 \\ k(4(x-.5)^{k-1} & .5 \le x < .75 \\ k(4(1-x))^{k-1} & .75 \le x < 1.0\,. \end{cases} \tag{6.3}$$

Hypothesis C's density has two peaks, one centered at $x = .25$ and the other at $x = .75$.

$$H_D(k):\ f(x) = \begin{cases} k(8x)^{k-1} & 0 \le x < .125 \\ k(8(.25-x))^{k-1} & .125 \le x < .25\,. \end{cases} \tag{6.4}$$

The pattern of this density is repeated for the intervals $.25 \le x < .5$, $.5 \le x < .75$, and $.75 \le x < 1.0$, resulting in four peaks when $k > 1$.

We considered the 1- through 4-point K-S tests and the 1- through 4-point Cramér-von Mises tests, with $k = 1.2, 1.5, 2.0, 3.0$, and 4.0. For each value of k

and for each alternative hypothesis B, C,and D, we generated 10 runs of $10,000$ experiments with sample size $n = 20$. All eight statistics were computed and compared to their $\alpha = .10$ and $\alpha = .05$ critical values. The numbers presented in Tables 3 and 4 show the average power over the 10 runs, at $\alpha = .10$ and $\alpha = .05$ respectively, of each of the eight statistics.

INSERT TABLE 3 AND TABLE 4 APPROXIMATELY HERE

As one would expect, the results of the power study for $\alpha = .05$ and $\alpha = .10$ are very similar. The K-S test and the Cramér-von Mises test are better than the other partition-based tests for the uniform alternatives. However, both the K-S test and the Cramér-von Mises test fare very poorly with the peaked alternatives. A possible explanation is that because the peaks are evenly distributed, the empirical and the theoretical distribution functions never get too far apart.

For the 1-peak alternative hypothesis, Kuiper's test, Watson's U_n^2, KS_4 and CvM_4 have the most power. This is in accord with our intuition: since each of these tests uses a statistic based on interval partitions, each test is able to separate effectively the peak from the rest of the distribution.

For the 2-peak and 4-peak alternative hypotheses, Kuiper's test and Watson's U_n^2 are no longer able to isolate the peaks very effectively; while the 4-point tests KS_4 and CvM_4 are capable of this. This would explain why the power of the 4-point tests dominates the power of the other tests in this situation.

7. CONCLUSIONS

We have introduced a framework for obtaining consistent generalizations of the Cramér-von Mises test on $\mathbf{R}^1$, and Watson's U_n^2 test on the line or on $\mathbf{S}^1$, based on classes of partitions invariant under the action of various groups of transformations. Our approach extends the work of Kuiper (1960) and Watson (1961), which in turn generalized the Kolmogorov-Smirnov (K-S) and Cramér-von Mises goodness-of-fit tests. We have also explored the first few of the natural series of distribution-free goodness-of-fit tests that result. As anticipated, the 4-point partition-based tests are more powerful when an alternative distribution with four peaks is considered. The 4-point tests are almost as powerful as the 2-point test with respect to a two-peak alternative.

The generalizations discussed in detail here are based on classes of partitions of $\mathbf{R}^1$ or $\mathbf{S}^1$ invariant under the action of the group of general coordinate transformations (diffeomorphisms) of the circle or the line. Tests that are not distribution-free, and not consistent against all alternatives, are associated with more limited symmetry groups, such as the group of translations of $\mathbf{R}^1$ or rotations of $\mathbf{S}^1$. Thus we have, overall, a way to unify a rather diverse collection of tests in a single framework. It is an approach that can suggest additional, powerful tests tailored to the known symmetries of a situation.

The framework generalizes further, to Euclidean spaces of higher dimension and to homogeneous spaces for various group actions. This is a subject of our ongoing research.

ACKNOWLEDGMENT

We thank the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) for travel support and hospitality during the course of this research.

REFERENCES

Ahmad, I. (1993). Modification of some goodness-of-fit statistics to yield asymptotically normal null distributions. *Biometrika* **80**, 466-472.

Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika* **55**, 343-354.

Barr, D. R. and Shudde, R. H. (1973). A note on Kuiper's V_n statistic. *Biometrika* **60**, 663.

Beran, R. J. (1968). Testing for uniformity on a compact homogeneous space. *Journal of Applied Probability* **5**, 177-195.

Beran, R. J. (1969). Asymptotic theory of a class of tests for uniformity of a circular distribution. *Ann. Math. Statist.***40**, 1196-1206.

Cressie, N. (1977) On some properties of the scan statistic on the circle and the line. *J. Appl. Prob.* **14**, 272-283.

Feltz, C. J. and Goldin, G. A. (1992). Generalization of the Kolmogorov-Smirnov goodness-of-fit test, using group invariance. *Journal of Nonparametric Statistics* **1**, 357-370. DIMACS Tech. Report No. 91-25.

Gine, E. M. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.***3**, 1243-1266.

Hermans, M. and Rasson, J. P. (1985). A new Sobolev test for uniformity on the circle. *Biometrika* **72**, 698-702.

Huffer, F. (1988). Some results concerning random arcs on the circle. *Journal of Applied Probability* **25**, 833-838.

Kokic, P. N. (1987). On tests of uniformity for randomly distributed arcs on a circle. *Australian Journal of Statistics* **29**, 179-187.

Kuiper, N. H. (1960). Tests concerning random points on a circle. *Proc. Kon. Ned. Akad. van Wet. A* **63**, 38-47.

Prentice, M. J. (1978). On invariant tests of uniformity for directions and orientations. *Ann. Statist.***6**, 169-176.

Rao, J. S. (1976). Some tests based on arc-lengths for the circle. *Sankhya Ser. B* **38**, 329-338.

Rothman, E. D. (1972). Tests for uniformity of a circular distribution. *Sankhya Ser. A* **34**, 23-32.

Rukhin, A. L. (1972). Some statistical decisions about distributions on a circle for large samples. *Sankhyā Ser. A* **34**, 243-250.

Stephens, M. A. (1963). The distribution of the goodness-of-fit statistic U_n^2, *Biometrika* **50**, 303-313.

Stephens, M. A. (1965). The goodness-of-fit statistic V_N: distribution and significance points, *Biometrika* **52**, 309-321.

Stephens, M. A. (1969). A goodness-of-fit statistic for the circle, with some comparisons, *Biometrika* **56**, 161-168.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48**, 109-114.

Watson, G. S. (1967a). Another test for the uniformity of a circular distribution. *Biometrika* **54**, 675-677.

Watson, G. S. (1967b). Some problems in the statistics of directions. *Bull. Int. Statist. Inst.,* cited in Beran (1969).

Watson, G. S. (1974). Optimal invariant tests for uniformity. In *Studies in Probability and Statistics*, Jerusalem Academic Press, 121-128.

Table 1: Critical Values for a 3-Point Extension of the Cramér-von Mises Test for Data in $\mathbf{R}^1$

	α = .10	.05	.025	.01
n = 2	0.107	0.120	0.136	0.155
2	0.05328	0.05976	0.06788	0.07761
3	0.05283	0.06277	0.07207	0.08385
4	0.05400	0.06365	0.07383	0.08704
5	0.05455	0.06471	0.07482	0.08808
6	0.05492	0.06519	0.07555	0.08927
7	0.05520	0.06574	0.07614	0.08996
8	0.05539	0.06597	0.07646	0.09072
9	0.05550	0.06621	0.07687	0.09098
10	0.05569	0.06644	0.07712	0.09161
11	0.05579	0.06654	0.07746	0.09199
12	0.05583	0.06675	0.07770	0.09223
13	0.05588	0.06677	0.07771	0.09241
14	0.05597	0.06691	0.07788	0.09251
15	0.05606	0.06695	0.07794	0.09273
16	0.05609	0.06708	0.07814	0.09282
17	0.05611	0.06712	0.07833	0.09295
18	0.05610	0.06722	0.07820	0.09297
19	0.05623	0.06737	0.07849	0.09332
20	0.05630	0.06740	0.07849	0.09331
25	0.05631	0.06741	0.07860	0.09364
30	0.05633	0.06751	0.07877	0.09382
35	0.05651	0.06778	0.07896	0.09402
40	0.05653	0.06783	0.07912	0.09419
45	0.05651	0.06773	0.07893	0.09408
50	0.05654	0.06778	0.07917	0.09432

Table 2: Critical Values for a 4-Point Extension of the Cramér-von Mises Test for Data in $\mathbf{R}^1$ or $\mathbf{S}^1$

	α = .10	.05	.025	.01
n = 2	0.01308	0.01473	0.01566	0.01625
3	0.01224	0.01462	0.01690	0.01939
4	0.01286	0.01492	0.01681	0.01957
5	0.01290	0.01503	0.01717	0.01992
6	0.01302	0.01519	0.01731	0.02010
7	0.01309	0.01527	0.01742	0.02028
8	0.01314	0.01534	0.01753	0.02034
9	0.01318	0.01539	0.01757	0.02045
10	0.01321	0.01545	0.01766	0.02056
11	0.01324	0.01548	0.01770	0.02058
12	0.01326	0.01550	0.01774	0.02068
13	0.01326	0.01556	0.01776	0.02072
14	0.01329	0.01556	0.01780	0.02075
15	0.01330	0.01557	0.01781	0.02075
16	0.01333	0.01561	0.01784	0.02080
17	0.01334	0.01562	0.01788	0.02085
18	0.01334	0.01564	0.01788	0.02085
19	0.01335	0.01565	0.01792	0.02089
20	0.01337	0.01567	0.01796	0.02091
25	0.01338	0.01569	0.01793	0.02091
30	0.01339	0.01571	0.01801	0.02100
35	0.01343	0.01577	0.01806	0.02106
40	0.01344	0.01579	0.01808	0.02107
45	0.01343	0.01578	0.01807	0.02104
50	0.01343	0.01577	0.01808	0.02110

Table 3: Power of Various Partition-Based Tests for $n = 20$ and $\alpha = .10$ (*)

(a) Alternative hypothesis $H_A(r,s)$: $f(x) = \begin{cases} r & 0 \le x < .5 \\ s & .5 \le x < 1.0 \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
r/s = 1.2	.1218	.1145	.1131	.1010	.1217	.1139	.1172	.1128
1.5	.2036	.1704	.1646	.1515	.2050	.1741	.1868	.1664
2.0	.3772	.2998	.2842	.2481	.3819	.3077	.3427	.2887

(b) Alt. hypothesis $H_B(k)$, 1 peak: $f(x) = \begin{cases} k(2x)^{k-1} & 0 \le x < .5 \\ k(2(1-x))^{k-1} & .5 \le x < 1.0 \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
k = 1.2	.0957	.1485	.0945	.1358	.0880	.1514	.1069	.1470
1.5	.1323	.3437	.1269	.2860	.1129	.3661	.1959	.3410
2.0	.2674	.7299	.2167	.6276	.2791	.7767	.4950	.7370
3.0	.6727	.9890	.4061	.9658	.8067	.9952	.9391	.9917
4.0	.9196	.9998	.5538	.9988	.9832	1.0000	.9976	.9999

(c) Alt. hypothesis $H_C(k)$, 2 peaks: $f(x) = \begin{cases} k(4x)^{k-1} & 0 \le x < .25 \\ k(4(.5-x))^{k-1} & .25 \le x < .5 \\ \textit{etc., to fill the} & \textit{interval.} \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
k = 1.2	.1050	.1130	.1069	.1280	.0986	.1069	.1048	.1196
1.5	.1246	.1653	.1592	.2492	.1055	.1404	.1490	.2119
2.0	.1663	.2965	.3360	.5665	.1219	.2423	.2891	.4818
3.0	.2564	.5998	.7575	.9493	.1671	.5721	.6968	.9150
4.0	.3327	.8201	.9551	.9979	.2168	.8616	.9371	.9952

(d) Alt. hypothesis $H_D(k)$, 4 peaks: $f(x) = \begin{cases} k(8x)^{k-1} & 0 \le x < .125 \\ k(8(.25-x))^{k-1} & .125 \le x < .25 \\ \textit{etc., to fill the} & \textit{interval.} \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
k = 1.2	.1036	.1054	.1048	.1112	.0990	.1010	.1008	.1043
1.5	.1114	.1223	.1271	.1487	.1019	.1088	.1134	.1242
2.0	.1265	.1568	.1819	.2496	.1052	.1267	.1390	.1728
3.0	.1600	.2448	.3238	.4899	.1145	.1694	.2064	.3009
4.0	.1855	.3217	.4519	.6787	.1237	.2100	.2724	.4298

(*) Here KS_1 is the usual Kolmogorov-Smirnov test, KS_2 is Kuiper's test, CvM_1 is the usual Cramér-von Mises test, and CvM_2 is Watson's test. The others are new.

Table 4: Power of Various Partition-Based Tests for $n = 20$ and $\alpha = .05$ (*)

(a) Alternative hypothesis $H_A(r,s) : f(x) = \begin{cases} r & 0 \le x < .5 \\ s & .5 \le x < 1.0 \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
$r/s = 1.2$	.0652	.0597	.0585	.0557	.0641	.0596	.0624	.0587
1.5	.1245	.0981	.0941	.0837	.1233	.1006	.1128	.0951
2.0	.2654	.1950	.1825	.1532	.2639	.2023	.2376	.1879

(b) Alt. hypothesis $H_B(k)$, 1 peak: $f(x) = \begin{cases} k(2x)^{k-1} & 0 \le x < .5 \\ k(2(1-x))^{k-1} & .5 \le x < 1.0 \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
$k = 1.2$	.0454	.0824	.0463	.0739	.0385	.0845	.0523	.0806
1.5	.0590	.2287	.0651	.1787	.0414	.2478	.1081	.2261
2.0	.1247	.5980	.1248	.4825	.0978	.6555	.3478	.6097
3.0	.4112	.9698	.2603	.9237	.5090	.9860	.8737	.9771
4.0	.7370	.9993	.3654	.9957	.8871	.9999	.9918	.9997

(c) Alt. hypothesis $H_C(k)$, 2 peaks: $f(x) = \begin{cases} k(4x)^{k-1} & 0 \le x < .25 \\ k(4(.5-x))^{k-1} & .25 \le x < .5 \\ \textit{etc., to fill the} & \textit{interval.} \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
$k = 1.2$	.0537	.0575	.0532	.0678	.0490	.0543	.0537	.0613
1.5	.0690	.0910	.0854	.1521	.0521	.0739	.0781	.1187
2.0	.0966	.1776	.2047	.4249	.0610	.1304	.1624	.3187
3.0	.1583	.4166	.5939	.8942	.0852	.3373	.4859	.8113
4.0	.2145	.6406	.8830	.9928	.1092	.6248	.8139	.9796

(d) Alt. hypothesis $H_D(k)$, 4 peaks: $f(x) = \begin{cases} k(8x)^{k-1} & 0 \le x < .125 \\ k(8(.25-x))^{k-1} & .125 \le x < .25 \\ \textit{etc., to fill the} & \textit{interval.} \end{cases}$

	KS_1	KS_2	KS_3	KS_4	CvM_1	CvM_2	CvM_3	CvM_4
$k = 1.2$	.0516	.0526	.0531	.0566	.0490	.0509	.0511	.0523
1.5	.0560	.0635	.0666	.0791	.0503	.0552	.0582	.0640
2.0	.0646	.0856	.1040	.1501	.0521	.0649	.0729	.0937
3.0	.0834	.1399	.2020	.3412	.0571	.0895	.1139	.1760
4.0	.1023	.1955	.3036	.5230	.0611	.1130	.1553	.2662

(*) Here KS_1 is the usual Kolmogorov-Smirnov test, KS_2 is Kuiper's test, CvM_1 is the usual Cramér-von Mises test, and CvM_2 is Watson's test. The others are new.