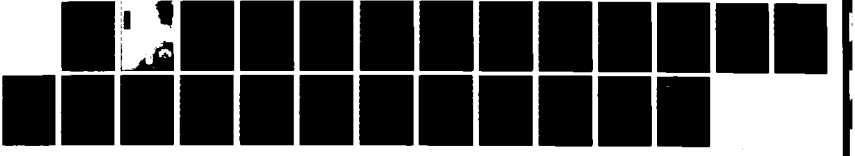
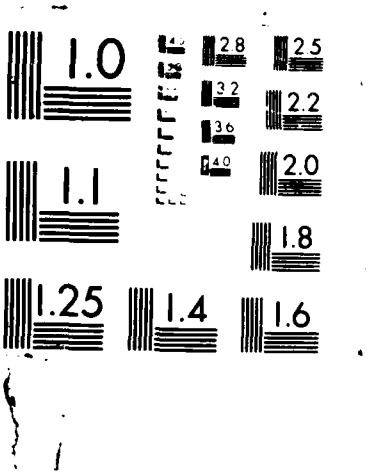
AD-A189 069 TESTS FOR UNIFORMITY ARISING FROM A SERIES OF EVENTS 1/1
(U) STANFORD UNIV CA DEPT OF STATISTICS M A STEPHENS
20 JAN 88 TR-400 N00014-86-K-0156

UNCLASSIFIED F/G 12/3 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

AD-A189

TESTS FOR UNIFORMITY ARISING FROM A SERIES OF EVENTS

BY

MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 400
JANUARY 28, 1988

Prepared Under Contract
N00014-86-K-0156 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TESTS FOR UNIFORMITY ARISING FROM A SERIES OF EVENTS

1. INTRODUCTION.

A number of problems in statistics reduce to testing whether a set of values comes from a uniform distribution: when the limits of this distribution are A,B we write it as $U(A,B)$. Most tests reduce to a test for $U(0,1)$. We concentrate here on test situations which arise from examining the times at which events occur. The scientist records these times on the time-axis: how should they then be analyzed?

If the events are governed by a Poisson process (that is, are occurring randomly in time), the intervals between events should have an exponential distribution and should be independent. If not, certain alternatives are often of interest. We therefore pose four questions which arise:

Question 1. Are the time intervals exponential, or

Question 2. Are the time intervals too regular or variable to be exponential?

Question 3. Are the intervals independent?

Question 4. Is a cluster of events (or more than one) present?

Formally, these problems can be set up in relation to the times of events, as follows. Let $0 < T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the times of observed events, occurring with continuous observation in the interval $(0,T)$; let $U_{(i)} = T_{(i)}/T$, $i=1,\ldots,n$, and define $U_{(0)} \equiv 0$ and $U_{(n+1)} \equiv 1$. Let the intervals between events be $E_i = T_{(i)} - T_{(i-1)}$, for $i=1,\ldots,n+1$, with $T_{(0)} \equiv 0$ and $T_{(n+1)} = T$; these lead to spacings $D_i = U_{(i)} - U_{(i-1)}$, $i=1,\ldots,n+1$, between the values $U_{(i)}$, and $D_i = E_i/T$.

Question 1 poses the question whether or not the E_i come from the exponential distribution $F(x) = 1-\exp(-x/\beta)$, $x > 0$; when they do and when they are independent, the transformation above gives a set $U_{(i)}$ which are ordered uniforms from $U(0,1)$. Thus the test for the Poisson process becomes a test of uniformity, specifically, a test of

$$H_0: \text{ the } U_{(i)} \text{ are ordered uniforms from } U(0,1).$$

The transformation from $T_{(i)}$ to $U_{(i)}$ effectively eliminates the unknown parameter β, which is of course connected with the unknown Poisson process rate λ.

When the times $T_{(i)}$ are given, it is very natural to examine the four questions above by transforming to the $U_{(i)}$ and then making a test of H_0. In this article we discuss, in this context, many of the possible tests available, and note some features which should be emphasized and on which further work is needed. For example, for some realistic alternatives some test statistics might have to be used with the lower tail whereas use of the upper tail is customary, thus changing accepted power comparisons, or, for others, existing results on power must be discounted because they do not apply to the practical alternatives. It may even be, of course, that the best tests of questions 1 to 4 are not made at all by transforming to the $U_{(i)}$ and then testing H_0.

This article forms part of a volume in honor of Professor Herbert Solomon, and among the test statistics are some on which Professor Solomon and I have worked together. These concern Neyman's tests and tests on spacings. Perhaps we should also note that many of the distributional properties of tests for uniformity involve elegant geometric probability arguments:

Professor Solomon has maintained a long interest in this field (Solomon, 1978), and it was my work on the null distribution of the goodness-of-fit statistic U^2, using arguments of geometric probability, which attracted his attention and so began our association many years ago.

2. ALTERNATIVES TO UNIFORMITY.

We first consider some situations of interest in connection with Questions 1 and 2.

(a) Testing That Lifetimes Are Exponential. Suppose the times $T_{(i)}$ are breakdown times for a machine because a part has failed; at each breakdown the part is immediately replaced, so that the intervals E_i are the lifetimes of the part.

It is common in reliability theory to test that such lifetimes are exponential, and, assuming they are independent, the resulting $T_{(i)}$ will be a realization of a Poisson process. The lifetime distributions, alternative to the exponential, which one might wish to detect, are then often the Weibull distribution

$$F_W(x) = 1-\exp-(\frac{x}{\beta})^{\gamma}, \quad x > 0$$

and the Gamma distribution $F_G(x) = \int_0^x f(t)dt$, where

$$f(t) = \frac{1}{\Gamma(\gamma)} t^{\gamma-1} e^{-t/\beta}, \quad t > 0.$$

If the E_i are from $F_G(x)$ or $F_W(x)$, we describe the resulting $D_i = E_i/T$ as scaled Gamma or scaled Weibull spacings; it will be important

that a test for uniformity of the $U_{(i)}$ should be able to detect scaled Gamma or scaled Weibull spacings. When $\gamma > 1$, in both Weibull and Gamma distributions, the lifetimes will be less spread out, for a given mean, than if they were exponential (the coefficient of variation is < 1, whereas for exponential it is 1); we can call these lifetimes super-regular. Super-regular lifetimes will produce a set $U_{(i)}$ which are excessively evenly spaced, more so than expected for a uniform sample. Stephens (1986b) calls such $U_{(i)}$ superuniform, and tests of H_0, in this context, should be able to detect superuniform $U_{(i)}$. Superregular lifetimes might be expected to occur, for example, if there is a high level of quality control for the machine component. If $\gamma < 1$, the lifetimes are super-variable, and lead to supervariable spacings between the $U_{(i)}$.

In reliability theory an important feature of a distribution is the failure rate, or hazard rate, given by $b(x) = f(x)/\{1-F(x)\}$. A distribution with decreasing failure rate (a DFR distribution) is such that $b(x)$ decreases as x increases, and for an increasing failure rate (IFR) distribution, $b(x)$ increases with x. An exponential distribution has constant failure rate $1/\beta$ for all values of x; Gamma and Weibull distributions $F_G(x)$ and $F_W(x)$ have DFR if $\gamma < 1$, and IFR if $\gamma > 1$. If lifetimes E_i come from, say, a DFR Weibull distribution, the spacings between smaller lifetimes are stochastically smaller than corresponding exponential lifetimes (that is, for the same n and i) and spacings between larger lifetimes are larger than corresponding exponential spacings. This cannot easily be detected, in a sequence of events, by looking at the lifetimes as naturally indexed by time; the lifetimes E_i would first have to be ordered by size, and the spacings between these compared with corresponding exponential spacings. Even this is not easily done, since one cannot easily carry

a mental picture of the relative sizes of exponential spacings. However, the following transformation can be used. Let $m = n+1$; we have m lifetimes E_i, including the last (unfinished) lifetime $T-T_n$. Let $E_i' = (m+1-i)(E_{(i)}-E_{(i-1)})$, where $E_{(i)}$ are the ordered E_i. $(i=1,\ldots,m;\ E_{(0)} \equiv 0)$. If the original E_i were exponentials, the E_i' will be unordered independent exponentials with the same scale parameter θ; but if the E_i were DFR, the E_i' will, on average, be increasing with i. Correspondingly reversed results hold for lifetimes from an increasing failure rate (IFR) distribution. Note that when the E_i are independent exponential, giving rise to independent exponential E_i', these intervals can be used to construct "times" $T'_{(1)} = E_1'$, $T'_{(2)} = E_1'+E_2'$, $T'_{(3)} = E_1'+E_2'+E_3'$,... etc., and, by scaling to give $U'_{(i)} = T'_{(i)}/T'_{(n+1)}$; the $U'_{(i)}$, $i=1,\ldots,n$, should be (ordered) $U(0,1)$. Tests based on the $U'_{(i)}$ are often used to test for exponentiality of the original E_i; however, we have now moved away from the naturally occurring time sequence $T_{(i)}$ and we shall not consider these tests at present: see Section 4.

(b) Testing That Lifetimes are Exponential, but with Average Lifetime Changing with Time. Again let the times $T_{(i)}$ be breakdown times, and, for a useful illustration, suppose they are mostly quite far apart. The lifetime of the replacement component might remain exponential as time passes, but, perhaps for instance because of improved manufacturing methods, the average lifetime (θ) increases with time. On the whole, then, the E_i are gradually becoming longer as time goes on; the process generating the $T_{(i)}$ can be viewed as Poisson but with rate λ no longer constant, but taking value $\lambda(t)$ varying with time. One would then expect the $T_{(i)}$, in this example, to be closer together at the left end (zero) than at the right end (one).

It might be worth observing that the distinction between the two situations, (a) and (b) above, can easily be blurred by conventional terminology. In (b), the machine breaks down less often if the component has increased lifetime, and the machine might then be colloquially described as having a decreasing failure rate (DFR). However, this is not the conventional use of this phrase in the theory of reliability.

In the lifetime model (b) just discussed, the exponential distribution is not fixed, but is changing as time goes by; β is increasing, and hence the spacings E_i are apparently getting longer. A graph of E_i against i would be increasing, whereas if β remained constant the values should hover around the horizontal line $E = \beta$. Also, if the E_i came from a fixed Weibull or Gamma distribution, the E_i, plotted against i, would again be horizontal around the mean value of the distribution. Tests might be based on such graphs: see Stephens (1986b) for further comment.

To sum up this discussion, we might detect Gamma or Weibull alternatives by looking for supervariable or superregular spacings between the $U_{(i)}$; or, when the spacings are superregular, by looking directly for superuniform $U_{(i)}$ themselves. If the lifetimes are exponential but with changing β, we must look for a drift of the $U_{(i)}$ towards one end or the other. However, other methods may be better than any of these, as we note in Section 4 below.

Of course, intervals can be superregular or supervariable without coming from Gamma or Weibull distributions, and a distributional test is not of primary interest. The events might be, for example, earthquakes, eruptions of volcanoes, or signals from a "black hole". If these are not Poisson, two important alternatives are that they are occurring at fairly

regular intervals, or perhaps the opposite; that is, the E_i are super-regular, or supervariable (Question 2 above). Earthquakes may occur in clusters, with several aftershocks and then long intervals between the clusters; the E_i would then appear supervariable. Regular intervals might be detected by eye, but overly dispersed intervals are much less easy to observe.

Questions 3 and 4. It might also be of interest, in observing the T_i, to detect a cluster, that is, a bunch of values too close to be represented by the same Poisson process as the others; for example, superimposed on random signals from the black hole, the friends of E.T. are sending frequent messages for him to call home. Detecting one or more important clusters will be very similar to detecting if the overall interval pattern shows too much dispersion.

Finally, there might be situations in which the successive intervals E_i are correlated, for example, a large one might tend to be followed by a small one. This could, for example, be the case when the E_i represents lengths of reigns of monarchs, where a long reign is followed by a short one because the heir is already older; in fact this may explain why the dates of reigns of English monarchs appear to be superuniform (Pearson, 1963).

3. TEST STATISTICS AND THE VARIOUS ALTERNATIVES.

In this section we examine how well test statistics for uniformity might be expected to detect the alternatives discussed in Section 2. The given set of times $T_{(i)}$ will be assumed to be converted to uniforms $U_{(i)}$ by $U_{(i)} = T_{(i)}/T$, i=1,...,n, as described above. (If T_n is not

known, we can take $T = T_{(n)}$, and have only n-1 uniforms $U_{(i)}$, since $U_{(n)} \equiv 1$.) Test statistics for the null hypothesis H_0: the $U_{(i)}$ are ordered uniforms from U(0,1), can be roughly classified into four families: (1) the Pearson X^2 statistic; (2) Neyman smooth tests; (3) EDF tests; (4) tests based on spacings. Comments follow on each of these families, and we shall finally concentrate on family (4).

(1) Pearson's X^2 statistic. This requires that the $U_{(i)}$ be classified into groups, preferably groups of equal probability; thus the line (0,1) is divided into k cells, and if N_j is the number of $U_{(i)}$ falling into the j-th cell, Pearson's statistic is $X^2 = k \sum_{j=1}^{k} (N_j - 1/k)^2$ with asymptotically a χ^2_{k-1} distribution. The grouping into cells loses much of the information in the $U_{(i)}$, especially for a small sample, and the X^2-statistic has low power against most alternatives. (Stephens, 1974; Quesenberry and Miller, 1977). Also, as usually used, large values of X^2 lead to rejection of H_0; X^2 will not then detect super-regularity of intervals unless small values are declared significant also.

(2) Neyman Smooth Tests. Neyman suggested that an alternative density to uniformity for U could be written

$$f(x) = c \exp\{1 + \sum_{j=1}^{k} \theta_j \ell_j(x)\}, \quad 0 < x < 1, \quad k = 1,2 \ .$$

where $\ell_1(x), \ell_2(x), \ldots$ are Legendre polynomials, $\theta_1, \theta_2, \ldots$ are parameters, and c, a function of the θ_j, is a normalizing constant. The Legendre polynomials are orthogonal on (0,1) and, by varying k, f(x) may be made to approximate any given alternative as closely as desired. Uniformity requires that all $\theta_j = 0$; thus the test H_0 can be put as a test that

$\Sigma_{j=1}^{k} \theta_j^2 = 0$. By likelihood ratio methods, Neyman formed the test statistic $N_k = \Sigma_{j=1}^{k} v_j^2$, where v_j is a Neyman component dependent on $\Sigma_{i=1}^{n} \xi_j(x_i)$. The interesting point of this method, in the present context, is that the first two components are functions of $\bar{U}$ and of the variance $s^2(U) = \Sigma(U_i - 0.5)^2/n$. A significant value for $\bar{U}$ might well occur in connection with Question 1, if the lifetimes were exponential, but θ was changing in time; and a significant value for $s^2(U)$ might arise if there is a cluster of events, or possibly negatively auto-correlated intervals, both tending to give a small variance of the U values. Percentage points for these statistics are given by Stephens (1966). (Note that it is the variance of the spacings D_i which would be required to examine Question 2 — see Greenwood's statistic below.) The individual components v_j are normalized to be $N(0,1)$ asymptotically; they are then also independent, so that for large n, $N_k^2 \sim \chi_k^2$. Some studies have indicated that N_2 is a good test statistic for general alternatives; the addition of further components can often weaken the overall power of N_k^2. However, N_k^2 approaches its asymptotic distribution χ_k^2 only slowly: Solomon and Stephens (1983) have recently given percentage points of N_2 based on fitting Pearson curves. Curve-fitting using moments, to obtain percentage points, has been one of our research interests in recent years (Solomon and Stephens, 1978), since computers have made such techniques much more practicable. Further comments on Neyman's statistics are in Solomon and Stephens (1983).

(3) EDF Tests. Tests based on the empirical distribution function are becoming increasingly well-known. The most famous of these is D, the Kolmogorov statistic; it is computed from $D^- = \max_i (U_{(i)} - (i-1)/n)$ and

$D^+ = \max_i(i/n - U_{(i)})$; $D = \max(D^+, D^-)$. Other statistics are $V = D^+ + D^-$; $W^2 = \sum U_{(i)} - (2i-1)/(2n)^2 + 1/(12n)$; $U^2 = W^2 - n(\bar{U}-1/2)^2$, where $\bar{U} = \Sigma U_{(i)}/n$; and

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log U_{(i)} + \log(1-U_{(n+1-i)})] + n \ .$$

As usually used, all these statistics are significant with large values only. (Note that U^2 is a statistic calculated from the $U_{(i)}$ of the sample; this terminology makes an unfortunate double use of U, but U^2 is the usual name for this statistic so we retain it.)

Statistic D^+ can be expected to detect a drift of $U_{(i)}$ toward 0, D^- a drift toward 1; however, if the direction of drift is not known, D must be used, and then it will often be less powerful than W^2 and, more particularly, A^2. As usually used, the statistics will not detect excessive regularity in the intervals - this would produce small values of the statistics, so lower tail tests must be considered. Autocorrelation will tend also to produce low values of these statistics. A cluster might be detected by U^2 or V, with large values significant, since these detect change in variance of the U-set. Further comparisons of tests for these situations need to be made.

(4) Spacings. The final group of tests to be considered here is the group based on the spacings $D_i = E_i/T$. Note that because the E_i are divided by their total, the D_i are not independent ($\sum_{i=1}^{n+1} D_i = 1$), nor are they distributed exponentially on H_0: the marginal distribution of any one spacing is $F(x) = 1-(1-x)^n$, $0 < x < 1$. Greenwood (1946) was one of the first to propose a test statistic for testing for randomness [illegible]

process) based on the D_i: Greenwood's statistic is $G = \Sigma D_i^2$. This statistic was investigated by Moran (1947,1953,1981); interest has revived again in recent years, and percentage points for finite n have been given by Burrows (1979), Currie (1981) and Stephens (1981). There has also been a great deal of interest in other functions of the D_i (see Pyke, 1965), and also in test statistics based on k-spacings. A k-spacing, for k fixed, is $D_{i,k} = U_{(i+k)} - U_{(i)}$, i=0,1,...,n+1-k; clearly D_i above is $D_{i,1}$ and for this special case k = 1 the second subscript will be omitted. Also $D_{i,k}$ is the sum of adjacent D_i, and as i varies, the $D_{i,k}$ overlap: non-overlapping spacings will be defined by $D^*_{i,k} = U_{(i+k)} - U_{(i)}$, i = 0,k,2k, etc. The k-spacings are sometimes called gaps or stretches. Much interesting work on the properties of k-spacings and statistics based on these has been done in recent years; for references see Stephens (1986a). Our interest here will be to see how this work relates to the questions posed above. Major families of test statistics based on spacings are:

(a) Greenwood's G and its extensions. The natural extension for G is $G_k = \sum_i D^2_{i,k}$ for fixed k = 1,2,3, etc, with the sum defined over the possible i-values, (G is then G_1) or $G^*_k = \sum_i D^{*2}_{i,k}$.

(b) $L_k = -\sum_i \log_e D_{i,k}$ or $L^*_k = -\sum_i \log_e D^*_{i,k}$.

(c) $M^+_k = \max_i D_{i,k}$ and $M^-_k = \min_i D_{i,k}$ for given k.

Statistic L_1 was first suggested by Moran (1951) in connection with testing for randomness of events. L_1 is the Maximum Likelihood test statistic for exponentiality of intervals against Gamma alternatives, so we might expect high power tests of Question 1 against this alternative.

Also, G is essentially the variance of the spacings: n times the variance is $(n-1)(n+1)$, so G might be useful in discussing Question 2. Of the many results concerning spacings statistics we highlight the following.

(a) Percentage points for G $(\equiv G_1)$ have been referred to above. Points for G_2 and G_3, and also for L_1, for finite n, have recently been given by McLaren and Stephens (1985). These were based on curve-fitting. G_k converges to asymptotic normality, but only very slowly: L_1 converges also slowly, to a distribution approximated by a χ^2 distribution.

(b) Suppose the alternative to uniformity is $f(x) = 1+t(x)/n^{1/2}$. Cibisov (1961) then showed the asymptotic relative efficiency (ARE) of tests based on spacings to be zero compared to EDF tests. However, Weiss (1965) gives an alternative which reverses this result. Among spacings themselves, Cressie (1979) showed, essentially, that tests using $D_{i,k}$ were asymptotically better than those using $D^*_{i,k}$ for alternatives which were step-functions, tending to the uniform like $n^{-1/4}$, slower than Cibisov's, and among such tests G_k was better than L_k using ARE as criterion. However, this cannot be the whole story: we have already observed that L_1 is the likelihood ratio test statistic in testing the spacings versus scaled Gamma spacing alternatives, and should have some optimal properties. The explanation appears to lie in the fact that the scaled Gamma spacings alternatives do not give a density for U which is either on the Cibisov or the Cressie model. Similar results appear to hold for Weibull spacings.

(c) McLaren and Stephens (1985) have investigated tests of H_0 against the alternative that the D_i are scaled Gamma spacings (Question 1 above and the Gamma alternative) and have found L_k better than G_k. McLaren

(1985) has calculated the ARE of G_k to L_k for this situation. The ARE A_k of G_k to L_k is only 0.39 when $k = 1$ (showing the expected excellent performance of L_1) but increases to $A_5 = 0.74$ and $A_{10} = 0.85$. From Monte Carlo studies for finite n, the L-group clearly dominates the G-group, and they both are better than EDF statistics. For detecting super-variable spacings (from a Gamma distribution with $\gamma < 1$), all these statistics are used in the "normal" way, with the upper tail significant. For detection of super regular intervals, from, for example, a Gamma distribution with $\gamma > 1$, which might well occur in practice, all the statistics must be used with the lower tail significant. A possible test statistic for super regular intervals would be M_k^+, looking to see if the largest k-spacing were too small.

In the above study on L_k and G_k, the powers declined considerably with increasing k. This seems reasonable; we know L_1 to be the Likelihood Ratio statistic, and unless the order in which the spacings appear is held to be important, there is no reason why a statistic which combines several in sequence should be especially effective.

(d) Autocorrelation. Quesenberry and Miller (1977) introduced the statistic

$$Q = \sum_{i=1}^{n+1} D_i^2 + \sum_{i=1}^{n} D_i D_{i+1}$$

to test for uniformity taking into account the possibility of autocorrelation between intervals. It is easy to see that $G_2 = 2Q - D_1^2 - D_{n+1}^2$, so that asymptotically Q is equivalent to G_2. McLaren and Stephens (1985) report a study on power of tests for autocorrelated intervals involving EDF statistics and also G_1, G_2, G_3 and L_1, L_2, L_3. The alternatives to uniform

spacings (themselves dependent because $\Sigma D_i = 1$) were autocorrelated scaled Gamma spacings. In this study, when correlation was positive, both the G_k and L_k had power increasing with k; G_2, G_3 and L_2, L_3 were better than all EDF statistics; L_2 and L_3 were best overall. All statistics were very poor at detecting negative autocorrelation such as might be expected in some practical problems where large intervals are compensated by small ones. Detection of this effect needs further study — in particular, the statistic $Q^* = \Sigma_{i=1}^{n} D_i D_{i+1}$ which forms part of Q might be effective standing alone.

(e) Searching for a Cluster. The EDF statistics U^2 and V, applied to the set U, will tend to detect a cluster, or a separation into two groups, one at each end of the (0,1) interval, since these statistics detect a shift in variance of the $U_{(i)}$ from the expected uniform value (Stephens, 1974) . Similarly the Neyman component v_2 will detect such a shift in variance. However, the presence of a cluster may not influence the overall variance enough to register significance with these statistics, and it is natural to look at M_k^- (to see if the minimum k-order gap is too small) to detect a cluster. Cressie (1977) has also examined the scan statistic — the maximum number of observations N_L, in a window of width L, as it moves along the (0,1) interval. Much work has been done, in particular by Nauss, Weiss, and more recently Cressie, on statistics M_k^- and N_L; see references in the papers cited in this subsection and in Stephens (1986a). As one might expect, there is a connection between the two statistics: $P(N_L > x) = P(M_x^- < L)$ (Nauss, 1966), so that a test based on one is equivalent to a test based on the other. Huntingdon and Nauss (1975) gave the exact distribution theory of M_k^-, but the formulas are

difficult; Cressie (1977) has given asymptotic theory, that, on H_0, $P(n^{(1+1/k)}M_k^- \leq x) = \exp(-x^k/k!)$ as $n \to \infty$ (recall that n is the number of uniforms $U_{(i)}$). Cressie has shown that the power of M_k^-, against a stepfunction alternative tending to the uniform like $n^{-1/4}$, is not as good as that of L_m or G_m. For other alternatives to clustering, M_k^- may have greater power, although it may be difficult to define these. The modelling of major earthquakes and aftershocks, where the aftershocks produce a cluster, would appear to be a possible practical application. A difficulty in applying these statistics is the choice of k, or the window width L; for α-levels to be correct, this must not be decided after looking at the times $T_{(i)}$ although this is a natural temptation. It would also be valuable if a sequential test were available, first seeing if M_1^- is too small, then M_2^-, etc.

A test for super-regularity of spacings might be based on M_k^+ (to see if the largest k-order gap is too small). Deken (1980) and Solomon and Stephens (1981) have given distribution theory and percentage points for M_k^+ for $n = 5$ and 10, and Deken gives also a Beta approximation for larger n. Similar remarks to those above apply to the choice of k, and to the desirability of a sequential test.

4. FINAL REMARKS.

In this article we have tried to draw attention to some of the outstanding questions which arise when tests on a series of events are converted to tests of uniformity. There are many ways in which events may depart from a random sequence, and this means that test statistics which are valuable for detecting one type of alternative will not be valuable for

another. The choice of test statistics in some situations is still an open question. Among factors to be considered are (a) some statistics may be significant in the tail not usually used; (b) some statistics may have a parameter which is difficult to choose (the order of Neyman's statistic, for example, or of a spacing or set of spacings); (c) existing power studies may not be applicable to the alternatives of practical concern.

Two other important issues should also be raised. One is that, with modern computer techniques, many statisticians will calculate many statistics and look at them all. Then the above factors, couched as they are in the classical language of hypothesis testing, will be less important: formal testing will not be applicable, since the final significance level is impossible to determine, and the best way to use the statistics is to allow them, or their significance levels, to throw light on the data in the knowledge of what different alternatives might be expected to give.

Another question to be considered is whether or not it is always useful to use the transformation to $U_{(i)}$, simple though it may be. It is persuasive that to discuss autocorrelation, or a change in exponential parameter, or clustering, one would examine the times and time intervals in situ; it is not so clear that to test that intervals are exponential with constant β, as opposed, say, to Gamma or Weibull, one should keep the intervals as they occur, and it may be best, for example, to look at them in order of size. The construction of E_i' and then $U'_{(i)}$, in Section 2(a), uses the size order of the intervals, and an extensive literature exists on tests based on the E_i' or the $U'_{(i)}$; they are related to the total time on test statistics and have much merit in terms of power (for some discussion see Stephens, 1986b).

Finally, we should not forget the wider question which often lies behind statistical examination of events – how to model the events realistically. In all the applications which have been alluded to here – the incidence of disease, lifetimes in reliability theory, signals from outer space, earthquakes, eruption of volcanoes, and of course in many others, a good model will suggest preferential statistical techniques. Some interesting comments on modelling, relevant to spacings statistics, are in the discussion to Pyke (1965) and the points made then are still pertinent twenty years later. This article merely attempts to see what different statistics might be expected to do for us, and to suggest, in addition, where work still must be done.

REFERENCES

1. Burrows, P.M. (1979). Selected percentage points of Greenwood's statistics. J. R. Statist. Soc. A, 142, 256-258.

2. Cibisov, D.M. (1961). On tests of fit based on sample spacings. Teor. Verovatnost. i Primenen., 6, 354-358.

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

4. Cressie, N. (1979). An optimal statistic based on higher order gaps. Biometrika, 66, 619-628.

5. Currie, I.D. (1981). Further percentage points of Greenwood's statistic. J. R. Statist. Soc. A, 144, 360-363.

6. Deken, J.G. (1980). Exact distributions for gaps and stretches. Technical Report, Department of Statistics, Stanford University.

7. Greenwood, M. (1946). The statistical study of infectious diseases. J. R. Statist. Soc. A, 109, 85-109.

8. Huntingdon, R.J. and Nauss, J.I. (1975). A simpler expression for k^{th} nearest neighbor coincidence probabilities. Ann. Prob., 3, 894-896.

9. McLaren, C.G. (1985). Contributions to Goodness of Fit. Ph.D. Thesis, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, B.C. Canada V5A 1S6.

10. McLaren, C.G. and Stephens, M.A. (1985). Tests for uniformity based on spacings. T.R., Dept. of Math and Stat., Simon Fraser University.

11. Moran, P.A.P. (1947). The random division of an interval. J. R. Statist Soc. B, 9. 92-98.

12. Moran, P.A.P. (1953). The random division of an interval-Part III. J. R. Statist. Soc. B, 15, 77-80.

13. Moran, P.A.P. (1981). Corrigendum: The random division of an interval. J. R. Statist. Soc. A, 144, 388.

14. Nauss, J.I. (1966). Some probabilities, expectations and variances for the size of largest clusters and smallest intervals. J. Amer. Statist. Assoc., 61, 1191-1199.

15. Pearson, E.S. (1963). Comparisons of tests for randomness of points on a line. Biometrika, 50, 315-326.

16. Pyke, R. (1965). Spacings. J. R. Statist. Soc. B, 27, 395-449.

17. Quesenberry, C.P. and Miller, F.L. (1977). Power studies of some tests for uniformity. J. Statist. Comput. and Simul., 5, 169-191.

18. Solomon, H. (1978). Geometric Probability SIAM App. Math. Series. Philadelphia: SIAM.

19. Solomon, H. and Stephens, M.A. (1978). The use of Pearson curves as approximations to densities. J. Amer. Statist. Assoc., 73, 153-160.

20. Solomon, H. and Stephens, M.A. (1981). On tests for uniformity: Neyman's statistic and statistics based on gaps and stretches. Technical Report 311, Dept. of Statistics, Stanford University.

21. Solomon, H. and Stephens, M.A. (1983). On Neyman's statistic for testing uniformity, Commun. Statist. Simula. Computa., 12, 127-134.

22. Stephens, M.A. (1974). EDF statistics for goodness-of-fit and some comparisons. J. Amer. Statist. Assoc., 69, 730-737.

23. Stephens, M.A. (1981). Further percentage points for Greenwood's statistic. J. R. Statist. Soc. A, 144, 364-366.

24. Stephens, M.A. (1986a). Tests for uniformity. Chap. 8 in Goodness-of-fit Techniques (eds. R.B. d'Agostino and M.A. Stephens) New York: Marcel Dekker.

25. Stephens, M.A. (1986b). Tests for exponentiality. Chap. 10 in Goodness-of Fit Techniques (eds. R.B. d'Agostino and M.A. Stephens). New York: Marcel Dekker.

26. Weiss, L. (1965). Contribution to the discussion of Pyke, R. (1965).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 400	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Tests For Uniformity Arising From A Series Of Events		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael A. Stephens		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE January 28, 1988 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Exponential distribution; renewal processes; Gamma distribution; Weibull distribution; Uniform distribution; tests for exponentiality; EDF tests; Neyman's test; spacings.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Test procedures are reviewed for testing whether events in time are occurring randomly. When this is so, a well-known transformation takes the times of occurrence to uniform order statistics, and the problem reduces to finding tests for uniformity which are sensitive to those alternative patterns of events which might be met in practice. We describe some alternatives, and discuss use of Neyman's test procedure, EDF tests, and spacings. Some recent work on spacings is examined in this context, and throughout the paper suggestions are made for future work.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE

FILMED

MARCH

1988

DTIC