The Discrete Runs Test and the Discrete Maximum of t Test

Technical Report CS 96-15

Andrew Shapira
ECSE Department
Rensselaer Polytechnic Institute
Troy, NY 12180

June 10, 1996

Abstract

Empirical tests are used to help determine whether a given sequence of numbers has statistical properties similar to those of a sequence whose elements are drawn from a uniform probability distribution. The runs test and the maximum of t test are two traditional tests that have been applied to sequences of real numbers between 0 and 1. Motivated by the desire for a set of empirical tests that all follow the pattern of taking n measurements and performing a chi-square test on the n measurements, this report examines discrete versions of both tests that are applied directly to any sequence whose elements are drawn from $\{0, 1, \ldots, d-1\}$, for some integer $d \geq 2$. The tests are practical and do not involve converting integers to a pseudo-continuous number through division by d. Expressions are derived for the probabilities needed to carry out the discrete tests. The utility of the discrete tests is compared to that of the continuous versions. When the tests are considered alone, and not within a collection of other tests, the discrete runs test seems preferable to the continuous version for all values of d, whereas the discrete maximum of t test appears to be preferable to the continuous version mainly for small values of d.

Key words. runs test, maximum of t test, random number testing, random number generation, chi-square test, empirical testing.

1 Introduction

Empirical tests are used to help determine whether a given sequence of numbers has statistical properties similar to those of a sequence whose elements are drawn from a uniform probability distribution. One application is the evaluation of pseudo-random number generators.

Two traditional tests are the runs test and the maximum of t test [Knu81]. These tests have been applied to sequences of real numbers between 0 and 1.

This report presents discrete versions of the runs test and the maximum of t test that are applied to sequences of integers drawn from $\{0, 1, \ldots, d-1\}$, for some integer $d \geq 2$. The tests are practical and do not involve converting integers to a pseudo-continuous number between 0 and 1 through division by d.

The discrete tests have some advantages over the continuous versions. A major advantage is that they help to create uniformity among tests. The discrete tests follow the common pattern of taking n measurements and performing a chi-square test on the n measurements. Test uniformity is an important factor in the complexity of a software implementation of a set of tests, and in interpreting test results.

Even if the issue of test uniformity is not considered, the discrete runs test appears to be preferable to the continuous version for all d. The discrete maximum of t test appears to be preferable to the continuous version mainly for small values of d. This is discussed in Section 6.

The discrete versions of these tests seem to have escaped notice until now. They are apparently not mentioned in discussions of the continuous versions of the tests in Knuth Volume 2 [Knu81] and Tezuka's book about uniform random numbers [Tez95]. Another suggestion that these tests are unknown may be found in one of L'Ecuyer's papers on pseudo-random number generation [L'E88]. There, results of roughly 20 empirical tests are reported for various pseudo-random generators. These tests are the continuous tests from [Knu81]; discrete versions of the tests are not mentioned.

An outline of the remainder of this report is as follows. Section 2 discusses the context in which the tests are run. Sections 3 and 4 define the tests, and derive the probabilities needed to carry them out. Section 5 describes how the tests are performed using the chi-square test. Section 6 briefly compares the discrete and continuous versions of the tests.

The following notation is used. For integer $d \geq 1$, we denote the set of integers $\{0, 1, \ldots, d-1\}$ by Z^d. The set of integers $\{a, a+1, \ldots, b\}$ is denoted $[a, b]$, for integers a and $b \geq a$.

2 Model of Tests

Our model of a test is as follows. Assume that we are given some sequence of numbers. A *measurement* is a quantity computed from designated sequence elements. In a *test*, one or more measurements are collected, and the collection of measurements is evaluated. This evaluation is usually made easier if the measurements are statistically independent.

Informally, in the runs test, a measurement is taken by gathering consecutive sequence elements until the gathered sequence is not monotonically increasing; the measured value is the length of the gathered sequence. In the maximum of t test, a measurement is taken by gathering t consecutive sequence elements; the measured value is the maximum value among the t elements.

The evaluation of a collection of measurements requires knowing the probability distribution of the possible measurement outcomes. These distributions are derived in Section 3.2 and Section 4 respectively for the discrete runs test and the discrete maximum of t test. Section 5 discusses using the standard chi-square statistical test with the discrete tests.

3 The Discrete Runs Test

3.1 Definition of the Test

In a measurement for discrete runs test, we are given a sequence of integers $X = X_1, X_2, \ldots,$ where $X_i \in Z^d$ for all X_i. The sequence X is assumed to contain at least one pair (X_i, X_{i+1}) such that $X_{i+1} < X_i$. The *run length* R_X of X is defined to be unique positive integer r such that

$$X_1 \leq X_2 \leq \cdots \leq X_r > X_{r+1}.$$

The measured value is the integer $\min(R_X, t)$, where $t \geq 2$ is a parameter to the test. If we are taking more than one measurement and we want the measurements to be independent, the next measurement uses the sequence $X_{R_X+2}, X_{R_X+3}, \ldots$, and not $X_{R_X+1}, X_{R_X+2}, \ldots$.

Using non-strictly increasing runs as above appears to provide a more powerful test than using strictly increasing runs. For example, when $d = 2$, requiring runs to be strictly increasing yields only two possible run lengths, whereas with non-strictly increasing runs, infinitely many run lengths are possible, and the space of possible results seems more interesting.

In an implementation of the test, it may be desirable to eventually give up if presented with a monotonically increasing sequence, rather than waiting indefinitely. We can introduce a parameter m to the test, and if a sequence's run length is found to exceed m, an error can be generated.

The above measurement is called a measurement for the *discrete runs up test with parameters d, t, and m*. A measurement for the *discrete runs down* test can defined similarly. The probabilities for the runs up test that are derived in the next section are the same as the probabilities for the runs down test.

3.2 Derivation of Probabilities

To carry out the discrete runs test, we want an expression for the probability $P(d,r) = P(R_Y = r)$, for arbitrary integers $d \geq 2$ and $r \geq 1$, when $Y = Y_1, Y_2, \ldots$ is a sequence with each Y_i drawn from the uniform distribution over Z^d. In this section we derive an expression for $P(d,r)$.

The idea is to count the number $T(r)$ of $(r+1)$-tuples whose elements are drawn from Z^d and constitute a run of length r. The desired probability is

$$P(d, r) = \frac{T(r)}{d^{r+1}}. \tag{1}$$

We will now find an expression for $T(r)$.

Given an integer $k \geq 0$, for all $r \geq 1$ we define $N(r, k)$ to be the number of r-tuples that are monotonically increasing and end in k:

$$N(r, k) = |\{(X_1, X_2, \dots, X_r) : 0 \leq X_1 \leq X_2 \leq \dots \leq X_r = k\}|.$$

A standard combinatorial device shows a one-to-one correspondence between the tuples that are counted by $N(r, k)$, and the selection of k objects from $r+k-1$ objects. This selection can be displayed graphically using an (r, k)-*configuration.* An (r, k) configuration is a string of $r + k - 1$ circles, exactly k of which are solid. For example, the string $\circ \bullet \circ \circ \circ \bullet \bullet \bullet \circ \bullet \bullet \bullet$ is the $(6, 7)$ configuration corresponding to the tuple $(0, 1, 1, 1, 4, 7)$.

In general, to convert an (r, k) configuration to an r-tuple, we read the configuration left to right, using an auxiliary variable i that is initialized to 0. When an empty circle is encountered, we print the current value of i; when a solid circle is encountered, we increment i and do not print anything. After encountering the rightmost circle we print k. Below we have shown the tuples that are counted by $N(3, 2)$, and the associated $(3, 2)$-configurations.

Tuple	Configuration
(0,0,2)	○ ○ ● ●
(0,1,2)	○ ● ○ ●
(0,2,2)	○ ● ● ○
(1,1,2)	● ○ ○ ●
(1,2,2)	● ○ ● ○
(2,2,2)	● ● ○ ○

Given k and a tuple of length r that is counted by $N(r,k)$, it should be clear how to generate a configuration that uniquely represents the tuple. Thus the one-to-one correspondence is established.

The number of (r,k) configurations is $\binom{r+k-1}{k}$. Therefore,

$$N(r,k) = \binom{r+k-1}{k}, \qquad r \geq 1, k \geq 0.$$

If $X_1, X_2, \ldots, X_r, X_{r+1}$ is a run up of length r, then $X_{r+1} \in Z_{X_r}$. Summing over the possible values of X_r, we have

$$\begin{aligned} T(r) &= \sum_{k=1}^{d-1} |Z^k| \cdot N(r,k) \\ &= \sum_{k=1}^{d-1} k \binom{r+k-1}{k}. \end{aligned}$$

We can rewrite the summand using symmetry, and then simplify:

$$\begin{aligned}
T(r) &= \sum_{k=1}^{d-1} r \binom{r+k-1}{r} \\
&= r \sum_{k=1}^{d-1} \binom{r+k-1}{r} \\
&= r \sum_{k=0}^{d-2} \binom{r+k}{r} \\
&= r \sum_{k=r}^{r+d-2} \binom{k}{r} \\
&= r \sum_{k=0}^{r+d-2} \binom{k}{r} \\
&= r \binom{r+d-1}{r+1}. \qquad (2)
\end{aligned}$$

The last step above follows from the "upper summation" formula [GKP89]. The derivation is completed by substituting (2) into (1), yielding

$$P(d,r) = \frac{r\binom{r+d-1}{r+1}}{d^{r+1}}.$$

Table 1 shows some of the discrete probabilities $P(d,r)$, and the probability $P(\infty,r)$. Table entries are shown rounded to 3 digits after the decimal point. Values less than .0005 have been omitted for clarity. The probability $P(\infty,r)$ is the probability from the continuous runs test, and is discussed in the next section.

3.3 Comparison with the Continuous Case

In the continuous runs test we have a sequence $V_1, V_2, \ldots$ of real numbers. When all V_i's are drawn from a uniform distribution over the same interval, the

d	r 1	2	3	4	5	6	7	8	9	10
2	.250	.250	.188	.125	.078	.047	.027	.016	.009	.005
4	.375	.313	.176	.082	.034	.013	.005	.002	.001	
8	.438	.328	.154	.056	.018	.005	.001			
16	.469	.332	.140	.044	.012	.003	.001			
32	.484	.333	.133	.039	.009	.002				
64	.492	.333	.129	.036	.008	.001				
128	.496	.333	.127	.035	.007	.001				
256	.498	.333	.126	.034	.007	.001				
512	.499	.333	.125	.034	.007	.001				
∞	.500	.333	.125	.033	.007	.001				

Table 1: Selected discrete runs test probabilities $P(d, r)$.

probability $P(\infty, r)$ of a run up of length $r \geq 1$ is

$$P(\infty, r) = \frac{r}{(r+1)!}. \tag{3}$$

For a given value of r, the discrete runs test probability $P(d, r)$ and the continuous runs test probability $P(\infty, r)$ become approximately equal as d increases.

3.4 The Runs Test Probability Distributions

For a given value of $d \geq 2$, a probability space is defined whose events are the possible run lengths (the positive integers), and each run length $r \geq 1$ occurs with probability equal to the discrete runs test probability $P(d, r) = \frac{r\binom{r+d-1}{r+1}}{d^{r+1}}$.

This probability distribution is similar to the negative binomial distribution, but is not identical to it. We were unable to transform this distribution into one of the standard probability distributions.

Since the probabilities in a probability space sum to 1, the discrete and continuous runs tests respectively yield the following identities:

$$\sum_{r\geq 1} \frac{r\binom{r+d-1}{r+1}}{d^{r+1}} = 1 \qquad \text{integer } d \geq 2, \tag{4}$$

$$\sum_{r\geq 1} \frac{r}{(r+1)!} = 1. \tag{5}$$

4 The Discrete Maximum of t Test

In a measurement for the maximum of t test, we are given a t-tuple $X = (X_1, X_2, \ldots, X_t)$, $t \geq 1$, with each $X_i \in Z^d$, $d \geq 2$. The measured value is the integer $M_X = \max_{1\leq i\leq t} X_i$.

To carry out the test, we want an expression for the probability $P_d(M_Y = k)$, for arbitrary $k \in Z^d$, when $Y = (Y_1, Y_2, \ldots, Y_t)$ is a tuple with each Y_i drawn from the uniform distribution over Z^d.

The probability $P_d(M_y = k)$ can be derived as follows. The Y_i's are mutually independent, so

$$\begin{aligned} P_d(M_Y \leq k) &= P_d(Y_1 \leq k) P_d(Y_2 \leq k) \cdots P_d(Y_t \leq k) \\ &= \left(\frac{k+1}{d}\right)^t. \end{aligned}$$

We have

$$P_d(M_Y = k) = P_d(M_Y \leq k) - P_d(M_Y \leq k-1).$$

Thus

$$P_d(M_Y = k) = \frac{(k+1)^t - k^t}{d^t}, \qquad 0 \leq k \leq d-1. \tag{6}$$

5 Using the Chi-Square Test

We can use a chi-square test to evaluate a collection of measurements [Coc52, Knu81]. In this section we briefly describe the chi-square test in general, and then look at issues specific to the discrete runs test and the maximum of t test.

Let P be a probability space where only finitely many events have nonzero probability. Denote the set of k events with nonzero probability by $Z^k = \{0, 1, \ldots, k-1\}$, and associate probability p_i with each event $i \in Z^k$. Each $i \in Z^k$ is called an *outcome.*

Now suppose we have n measurements whose possible outcomes are in Z^k, the measurements are independent and drawn from the same probability distribution, and this distribution is unknown to us. The set of measurements defines an observed frequency distribution; for each $i \in Z^k$, let f_i denote the number of measurements whose value is i.

In the chi-square test, we form the *chi-square statistic*

$$V = \sum_{i=0}^{k-1} \frac{(f_i - np_i)^2}{np_i}.$$

A *confidence value* between 0 and 1 is then associated with V, using software or a precomputed table. This confidence value is the probability that n samples drawn from P would produce a chi-square statistic less than or equal to V. If we obtain a confidence value that is extremely close to 0 or 1, then this may be taken as evidence that the samples were not drawn from P.

Confidence values in the chi-square test are usually calculated using approximations. A rule of thumb for making the error acceptably small is that each outcome in P should have an expected frequency of at least 5 [Knu81]. The person carrying out the test can pick a real number s, and adjust the number of samples n so that the expected frequency count for each outcome is at least s. Another approach is to lump outcomes together; this allows us to define a new probability space in which each event has an expected frequency count of at least s. When such lumping is used, we use the term *category* to denote the events in the new probability space. When lumping is not used, we use *category* and *outcome* interchangeably.

In the continuous runs test or the discrete runs test, we can pick an integer u, and form one of the categories by lumping together all outcomes that correspond to runs of length u or more. The other categories are formed by associating each run length smaller than u with a distinct category. We select n, u, and s so each category has an expected frequency of at least s. Given n and s, choosing the maximum possible value of u is straightforward. This procedure can be used for the continuous or discrete runs test.

The discrete maximum of t test is more complicated. We can lump outcomes together using a scheme like this: divide $[0, d-1]$ into intervals $[0, a_1]$, $[a_1+1, a_2]$, $\ldots$, $[a_{k-1}+1, d-1]$, so that the expected frequency of each interval is at least s. One chi-square category corresponds to each interval. If a t-tuple has a maximum value of M, then we can use a binary search to find the interval that

M lies in, and increment that interval's frequency count.

The most difficult task in this scheme is to determine the intervals themselves. For a given value of d, the probability associated with the interval $[i, j]$, $j \geq i$, is $P_d([i, j]) = \frac{(j+1)^t - i^t}{d^t}$. We can determine the intervals if we can solve the following problem: given an integer $i \in [0, d-1]$, find the value $u_i \in [i, d-1]$ (if it exists) such that $P_d([i, u_i]) \geq \frac{s}{n}$ and $P_d([i, u_i - 1]) < \frac{s}{n}$. The first interval is $[0, u_0]$, the second interval is $[u_0 + 1, u_{(u_0+1)}]$, and so on. If necessary, we can lump the last two intervals together so that all intervals have an expected frequency of at least s.

Given an integer $i \in [0, d-1]$, one way to find u_i is to calculate it directly. It is easy to show that

$$u_i = \left\lceil \left(\frac{sd^t}{n} + i^t \right)^{1/t} - 1 \right\rceil \tag{7}$$

(assuming u_i exists). A problem with calculating u_i directly is that it can involve computing the t^{th} root of a large number. A different approach is simply to find u_i using a binary search of the interval $[i, d-1]$.

6 Discussion

An advantage of the discrete tests over their continuous counterparts is that the discrete tests do not use the Kolmogorov-Smirnov test (KS test). The KS test is designed for continuous distributions. Using the KS test with a discrete distribution seems to invite complications or errors that may be avoidable through the use of a chi-square test.

The continuous runs test uses the approximation $P(\infty, r) \approx P(d, r)$. As we saw in Table 1, this approximation is not good for small values of d. Thus it is inadvisable to use the continuous runs test with small values of d. On the other hand, the discrete runs test uses the exact probability. Also, for large values of d, the discrete runs test should be as effective as the continuous test. No reason is evident for preferring the continuous runs test over the discrete runs test.

For small values of d and t, the discrete maximum of t test can have a distinct chi-square category for each value in $[0, d-1]$, and furthermore we can avoid the potential problems that may arise from using an extremely discretized distribution in the KS test. For large values of d, one category may represent many discrete maximum of t outcomes. The discrete test remains valid for large values of d, but here the continuous version of the test may be preferable, since the discretization of the distribution in the KS test is not severe, and there is no outcome lumping.

The probabilities for the discrete tests clarify the error in the approximation that occurs when we use the continuous test probabilities on a sequence of integers drawn from Z^d. The discrete test probabilities can be computed with no rounding error, using arbitrary precision rational arithmetic.

The discrete versions of the tests help to create uniformity among empirical tests. It is desirable to have all tests to follow the common pattern of taking

n measurements and then performing a chi-square test. We may also want to repeat this pattern N times and do a KS test on the set of N chi-square confidence values.

Uniformity of tests was the main factor that motivated this investigation of the discrete versions of the tests. This uniformity facilitated the implementation of the discrete runs test and the discrete maximum of t test for the Randpack sequence generation and testing library [ES96, Sha96].

7 Acknowledgement

The author thanks Úlfar Erlingsson for reading a draft of this report.

References

[Coc52] William G. Cochran. The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, 23:315–345, 1952.

[ES96] Úlfar Erlingsson and Andrew Shapira. Randpack facilities for testing pseudo-random sequences, 1996. To be submitted for publication.

[GKP89] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, Reading, Mass., 1989.

[Knu81] Donald E. Knuth. *The Art of Computer Programming*, volume 2 (Semi-Numerical Algorithms). Addison-Wesley, Reading, Massachusets, 1981.

[L'E88] Pierre L'Ecuyer. Eficient and portable combined random number generators. *Communications of the ACM*, 31(6):742–749,774, June 1988.

[Sha96] Andrew Shapira. Randpack facilities for generating pseudo-random sequences and presenting them to applications, 1996. To be submitted for publication.

[Tez95] Shu Tezuka. *Uniform Random Numbers: Theory and practice.* Kluwer Academic Publishers, 1995.