*Draft Version: March 14, 1997*
Submitted to *Journal of Statistical Computation and Simulation*

# Tests Based on Sum-Functions of Spacings for Uniform Random Numbers

PIERRE L'ECUYER
Département d'Informatique et de Recherche Opérationnelle (IRO)
Université de Montréal, C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7, Canada;
e-mail: lecuyer@iro.umontreal.ca
web: http://www.iro.umontreal.ca/~lecuyer

ABSTRACT: We examine the idea of testing uniform random number generators via two goodness-of-fit statistics: the sum of the logarithms and the sum of squares of overlapping $m$-spacings. The first statistic is related to an estimator of the entropy of a density and is good to detect clustering, whereas the second one, known as Greenwood's statistic for $m = 1$, is optimal in terms of Pitman efficiency, in certain setups, among sum-functions of $m$-spacings. These statistics are asymptotically normally distributed. We evaluate the distance between the standard normal distribution and that of the standardized statistics, as a function of $m$ and of the sample size, when standardization is done using either the asymptotic or the exact (for finite sample size) mean and variance. We then report on experiments with these statistics to detect defects in some popular random number generators.

**Key Words:** Random number generators; statistical tests; spacings.

## 1 Introduction

Let $U_1, \ldots, U_n$ be a sequence of independent random variables uniformly distributed over the interval $(0,1)$ (for short, i.i.d. $U(0,1)$). Let $U_{(1)} \le \ldots \le U_{(n)}$ be the $U_i$ sorted by increasing order. Define $U_{(0)} = 0$ and $U_{(n+1)} = 1$. The *overlapping spacings of order* $m$ (for $m < n$) are:

$$G_{m,i} = U_{(i+m)} - U_{(i)}, \qquad 0 \le i \le n - m + 1.$$

A general class of goodness-of-fit statistics is defined by

$$H_{m,n} = \sum_{i=0}^{n-m+1} h(nG_{m,i}) \tag{1}$$

where $h$ is a given "smooth" function (see, e.g., Pyke 1965, Kuo and Rao 1981, Rao and Kuo 1984, Hall 1986). There are several variants of this; for example, Hall (1986) replaces (1) by $\sum_{i=1}^{n-m} h(((n+1)/m)G_{m,i})$. The following *circular* version can be viewed as a function of $m$-spacings on a circle with unit circumference. Define $U_{(n+i)} = 1 + U_{(i-1)}$ for $1 < i \le m$, $G_{m,i}$ as above for $0 \le i \le n$, and

$$H_{m,n}^{(c)} = \sum_{i=0}^{n} h(nG_{m,i}). \tag{2}$$

Sum-functions of *disjoint* $m$-spacings, where the sum is only over $i = 0, m, 2m, \ldots$, have been considered by del Pino (1979) and others. Note that (1), (2), and their disjoint versions differ only for $m > 1$.

Under mild assumptions on $h$, all these statistics, properly standardized, are asymptotically normal as $n \to \infty$, either for fixed $m$ or when $m \to \infty$ with $m = O(n^p)$ for $0 < p < 1$ (see del Pino 1979, Hall 1986, Jammalamadaka, Zhou, and Tiwari 1989, Kuo and Rao 1981 and other references given there).

In this paper, we are interested in using these statistics to test the null hypothesis $H_0$ that a given sequence of numbers is a realization of a sequence of i.i.d. $U(0,1)$ random variables. Our aim is to apply this to assess the quality of uniform random number generators. For background on random number generation and statistical testing, see, for example, Hellekalek (1995), Knuth (1981), L'Ecuyer (1994), Marsaglia (1985), Niederreiter (1992). We consider the cases $h(x) = \ln(x)$ and $h(x) = x^2$. The latter yields the sum of squares of spacings, introduced by Greenwood (1946) for $m = 1$, and known to be the optimal choice of $h$ in terms of Pitman asymptotic efficiency, for a general class of alternatives that are "smooth" perturbations of the uniform, for fixed $m$ (Sethuraman and Rao 1970, Cressie 1979, Kuo and Rao 1981, Rao and Kuo 1984). The power also increases with $m$. The sum of logarithms of spacings has been studied at length (see, e.g., Cressie 1976, Cressie 1979, Pyke 1965) and is strongly related to an empirical entropy statistic recommended in Dudewicz and van der Meulen (1981) and Dudewicz et al. (1995). Other functions $h$ have been considered (e.g., Pyke 1965, Russell and Levitin 1995), but here, we limit ourselves to the two just mentioned.

A weakness of these spacings tests (in theory) is that for fixed $m$, they cannot distinguish alternatives densities at distance less than $O(n^{-1/4})$ from the uniform; i.e., for densities of the form

$$f_n(u) = 1 + n^{-\delta}\ell(u) \tag{3}$$

where $\ell$ is a continuously differentiable function on $[0, 1]$ such that $\int_0^1 \ell(u)du = 0$, the tests do not discriminate $f_n(u)$ from the uniform, asymptotically, if $\delta > 1/4$ (Sethuraman and Rao 1970, Kuo and Rao 1981). This is poor compared with other standard tests such as the Kolmogorov-Smirnov or Pearson's chi-square, who can discriminate similar alternatives at distance $O(n^{-1/2})$. In response to this, several authors, e.g., Cressie (1976), del Pino (1979), Dudewicz and van der Meulen (1981), suggested increasing $m$ to infinity, simultaneously with $n$. The main conclusions (Hall 1986, Jammalamadaka, Zhou, and Tiwari 1989) are that if $m/n \to 0$ as $n \to \infty$, then $H^{(c)}_{m,n}$ can discriminate alternatives of the form $f_n(u) = 1 + (mn)^{-1/4}\ell(u)$. If $m = n^p$ for $p < 1$, these alternatives are at distance $O(n^{-(1+p)/4})$, which can be made arbitrarily close to $O(n^{-1/2})$. This holds for any smooth enough $h$ and also for the disjoint spacings. If $m/n \to \gamma$ as $n \to \infty$, where $0 < \gamma < 1$, then $H^{(c)}_{m,n}$ discriminates alternatives at distance $O(n^{-1/2})$ if $\gamma$ is irrational and (strangely enough) not if $\gamma$ is rational. For $H_{m,n}$, however, alternatives at distance $O((mn)^{-1/4})$ are discriminated only if $m = O(n^{-1/2})$, due to edge effects.

Based on these properties, one would recommend $H^{(c)}_{m,n}$ with $m = n^p$, for $p$ close to 1. When testing random number generators, however, the alternative that corresponds to a generator's defect is usually *not* a smooth perturbation to the uniform density as in (3). For example, there could be clustering or repulsive tendencies between the points, or some complicated non-linear correlation, etc. Often, one does not have a specific class of alternatives in mind. So, it is not clear if the efficiency properties described above are relevant. Another issue is that convergence to the asymptotic normal distribution may take place at a much slower pace for larger $m$.

In this paper, we explore those issues. For $h(x) = \ln(x)$ and $h(x) = x^2$, we standardize $H_{m,n}^{(c)}$ by its *exact* mean and variance and estimate by simulation, for different $m$ and $n$, the Kolmogorov distance to the standard normal distribution. Because of better centering and scaling, this distance is significantly smaller than the corresponding distance when $H_{m,n}^{(c)}$ is standardized by the *asymptotic* mean and variance. In both cases, it increases with $m$. The circular version (2) generally behaves better than (1) in this respect when the asymptotic values are used for the mean and the variance, but not much when using the exact values. The sum of logarithms of spacings is equivalent, up to a linear transformation and except for minor edge effects, to the empirical entropy statistic proposed by Dudewicz and van der Meulen (1981). The latter is asymptotically standard normal, but even for large $n$, its distribution remains far from the standard normal compared with the sum of logarithms of spacings standardized by its exact mean and variance.

The next section summarizes some properties of the distribution of the sum of logarithms and the sum of squares of spacings. We derive expressions for the exact mean and variance of the sum of squares of spacings. The quality of approximation by the standard normal distribution is examined in Section 3. Section 4 discusses the effects of finite-precision arithmetic. In Section 5, we report on numerical experiments to illustrate the capability of these goodness-of-fit statistics to catch up defects in some popular random number generators. Certain linear congruential and nonlinear inversive generators fail the tests.

## 2 Normal approximation for sums-functions of spacings

Let $X_0, X_1, \ldots$ be i.i.d. exponential random variables with unit mean and $Y_{j,m} = X_j + \ldots + X_{j+m-1}$. Define $\mu_m = E[h(Y_{0,m})]$ and

$$\begin{aligned}\sigma_m^2 &= \mathrm{Var}[h(Y_{0,m})] + 2\sum_{j=1}^{m-1} \mathrm{Cov}[h(Y_{0,m}), h(Y_{j,m})] \\ &\quad -\frac{1}{m^2}\left( \mathrm{Cov}[Y_{0,m}, h(Y_{0,m})] + 2\sum_{j=1}^{m-1} \mathrm{Cov}[Y_{j,m}, h(Y_{0,m})]\right)^2 .\end{aligned}$$

Different versions of the following limit theorem appear in the literature. The version we give follows Hall (1984). Similar results hold when both $m$ and $n$ diverge simultaneously, as mentioned in the introduction.

PROPOSITION **1.** *Assume that $h$ is continuously differentiable on $(0,\infty)$ and that there exist positive constants $K$, $a$, $b$, and $\epsilon$ such that*

$$|h(x)| \le K(x^a + x^{-m+\epsilon}) \qquad \text{and} \qquad |h'(x)| \le K(x^b + x^{-m/2-1+\epsilon}) \tag{4}$$

*for all $x > 0$. Then, as $n \to \infty$,*

$$H_{m,n}^{(A)} \stackrel{\text{def}}{=} \frac{H_{m,n} - (n-m+2)\mu_m}{\sigma_m\sqrt{n}} \Rightarrow N(0,1) \tag{5}$$

*and*

$$H_{m,n}^{(A,c)} \stackrel{\text{def}}{=} \frac{H_{m,n}^{(c)} - (n+1)\mu_m}{\sigma_m\sqrt{n}} \Rightarrow N(0,1), \tag{6}$$

*where $\Rightarrow$ denotes convergence in distribution and $N(0,1)$ is the standard normal.*

To reduce the distance to the $N(0,1)$ distribution for finite $n$, one can replace the centering and scaling expressions in (5) and (6) by the exact mean and variance of $H_{m,n}$ or $H_{m,n}^{(c)}$. This gives

$$H_{m,n}^{(E)} \stackrel{\text{def}}{=} \frac{H_{m,n} - E[H_{m,n}]}{\sqrt{\operatorname{Var}[H_{m,n}]}} \tag{7}$$

and

$$H_{m,n}^{(E,c)} \stackrel{\text{def}}{=} \frac{H_{m,n}^{(c)} - E[H_{m,n}^{(c)}]}{\sqrt{\operatorname{Var}[H_{m,n}^{(c)}]}}. \tag{8}$$

These also converge in distribution to the $N(0,1)$. In the previous four equations, the $A$ and $E$ superscripts stand for *asymptotic* and *exact*, respectively.

From now on, we denote $H$ by $L$ when $h(x) = \ln(x)$ and by $S$ when $h(x) = x^2$. These functions satisfy the assumptions of Proposition 1. We thus have

$$L_{m,n} = \sum_{i=0}^{n-m+1} \ln(nG_{m,i}), \tag{9}$$

$$L_{m,n}^{(c)} = \sum_{i=0}^{n} \ln(nG_{m,i}), \tag{10}$$

$$S_{m,n} = \sum_{i=0}^{n-m+1} (nG_{m,i})^2, \tag{11}$$

$$S_{m,n}^{(c)} = \sum_{i=0}^{n} (nG_{m,i})^2. \tag{12}$$

The standardized versions are obtained by replacing $H$ by either $L$ or $S$ in the definitions of $H^{(A)}_{m,n}$, $H^{(A,c)}_{m,n}$, $H^{(E)}_{m,n}$, and $H^{(E,c)}_{m,n}$. For example,

$$S^{(E)}_{m,n} \stackrel{\text{def}}{=} \frac{S_{m,n} - E[S_{m,n}]}{\sqrt{\operatorname{Var}[S_{m,n}]}},$$

and so on.

The next proposition recalls expressions derived by Cressie (1976) for the exact mean and variance of the sums of logarithms of spacings (circular version). Its requires the following notation. For $1 \le j \le k \le \infty$, let

$$R(j,k) = \sum_{i=j}^{k} \frac{1}{i} \qquad \text{and} \qquad Q(j,k) = \sum_{i=j}^{k} \frac{1}{i^2}.$$

For $j > k$ define $R(j,k) = Q(j,k) = 0$. Recall that $R(1,k) - \ln(k+1) \to \gamma$ as $k \to \infty$, where $\gamma = 0.57721566...$ is Euler's constant, and that $Q(1,\infty) = \pi^2/6$.

Proposition **2.** *One has*

$$\begin{aligned} E[L^{(c)}_{m,n}] &= -(n+1)[R(m,n) - \ln(n)], \\ \operatorname{Var}[L^{(c)}_{m,n}] &= (n+1)\left[Q(m,n) + nQ(1,n) - 2(m-1)(mQ(1,m-1)+1) \right. \\ &\qquad \left. + (2m(m-1) - n)\pi^2/6\right]. \end{aligned}$$

*Proof.* Observe that $E[L^{(c)}_{m,n}] = (n+1)(\ln(n) + E[\ln(G_{m,0})])$ and $\operatorname{Var}[L^{(c)}_{m,n}] = \sum_{i=0}^{n}\sum_{j=0}^{n} \operatorname{Cov}[\ln(G_{m,i}), \ln(G_{m,j})]$. It then suffices to use the explicit expressions for the expectations and covariances obtained by Cressie (1976). □

We now derive similar results for the sum of squares of spacings, for both the linear and circular versions.

Proposition **3.** *One has*

$$E[S_{m,n}] = \frac{m(m+1)(n-m+2)n^2}{(n+1)(n+2)}, \tag{13}$$

$$E[S^{(c)}_{m,n}] = \frac{m(m+1)n^2}{n+2}, \tag{14}$$

$$\operatorname{Var}[S_{m,n}] = \frac{m(m+1)n^4}{3(n+4)(n+3)(n+2)^2(n+1)^2}\left[2n^3(2m+1) + n^2(8m^3 - 15m^2 + 3m + 16)\right.$$

$$+ n(-12m^4 + 48m^3 - 27m^2 - 37m + 34)$$
$$- 30m^4 + 58m^3 + 6m^2 - 54m + 20\Big] \qquad (15)$$

$$\mathrm{Var}[S^{(c)}_{m,n}] = \frac{2m(m+1)n^4[n(2m+1) - 3m^2 + m + 2]}{3(n+2)^2(n+3)(n+4)}. \qquad (16)$$

*Proof.* Recall (see Rényi 1953) that the order statistics $U_{(i)}$, $i = 1, \ldots, n$, can be represented as

$$U_{(i)} = \exp\left[-\frac{X_1}{n} - \frac{X_2}{n-1} - \cdots - \frac{X_{n-i+1}}{i}\right]$$

where $X_1, \ldots, X_n$ are i.i.d. exponentials with unit mean. With this representation,

$$G_{m,i} = \exp\left[-\frac{X_1}{n} - \cdots - \frac{X_{n-i-m+1}}{i+m}\right]\left(1 - \exp\left[-\frac{X_{n-i-m}}{i+m-1} - \cdots - \frac{X_{n-i+1}}{i}\right]\right)$$

for $1 \leq i \leq n$, while $G_{m,0} = U_{(m)}$. Then,

$$\begin{aligned} E[G^2_{m,0}] &= E\left[\exp\left[-\frac{2X_1}{n} - \frac{2X_2}{n-1} - \cdots - \frac{2X_{n-m+1}}{m}\right]\right] \\ &= \frac{n}{(n+2)}\frac{(n-1)}{(n+1)} \cdots \frac{(m+1)}{(m+3)}\frac{m}{(m+2)} \\ &= \frac{m(m+1)}{(n+1)(n+2)}. \end{aligned}$$

Since $E[S_{m,n}] = (n-m+2)n^2 E[G^2_{m,0}]$ and $E[S^{(c)}_{m,n}] = (n+1)n^2 E[G^2_{m,0}]$, (13) and (14) follow. We now prove (16). We have

$$\begin{aligned} \mathrm{Var}\left[S^{(c)}_{m,n}\right] &= n^4 \ \mathrm{Var}\left[\sum_{i=0}^{n} G^2_{m,i}\right] \\ &= n^4(n+1)\Bigg[\mathrm{Var}[G^2_{m,0}] + 2\sum_{j=1}^{m-1} \mathrm{Cov}[G^2_{m,0}G^2_{m,j}] \\ &\qquad + (n-2m+2)\ \mathrm{Cov}[G^2_{m,0}G^2_{m,m}]\Bigg]. \end{aligned} \qquad (17)$$

As we did for $E[G^2_{m,0}]$, we can write

$$\begin{aligned} E[G^4_{m,0}] &= E\left[\exp\left[-\frac{4X_1}{n} - \frac{4X_2}{n-1} - \cdots - \frac{4X_{n-m+1}}{m}\right]\right] \\ &= \frac{m(m+1)(m+2)(m+3)}{(n+1)(n+2)(n+3)(n+4)}. \end{aligned}$$

Also, for $0 < j \le m$,

$$\begin{aligned}
E[G_{m,0}^2 G_{m,j}^2] &= E\left[\exp\left[-\frac{2X_1}{n} - \cdots - \frac{2X_{n-m+1}}{m}\right] \exp\left[-\frac{2X_1}{n} - \cdots - \frac{2X_{n-m-j+1}}{m+j}\right]\right. \\
&\quad \left. \cdot \left(1 - \exp\left[-\frac{X_{n-i-m}}{i+m-1} - \cdots - \frac{X_{n-i+1}}{i}\right]\right)^2\right] \\
&= E\left[\exp\left[-\frac{4X_1}{n} - \cdots - \frac{4X_{n-m-j+1}}{m+j} - \frac{2X_{n-m-j}}{m+j-1} - \cdots - \frac{2X_{n-m+1}}{m}\right]\right. \\
&\quad + \exp\left[-\frac{4X_1}{n} - \cdots - \frac{4X_{n-m+1}}{m} - \frac{2X_{n-m}}{m-1} - \cdots - \frac{2X_{n-j+1}}{j}\right] \\
&\quad - 2\exp\left[-\frac{4X_1}{n} - \cdots - \frac{4X_{n-m-j+1}}{m+j} - \frac{3X_{n-m-j}}{m+j-1} - \cdots - \frac{3X_{n-m+1}}{m}\right. \\
&\quad \left.\left. - \frac{X_{n-m}}{m-1} - \cdots - \frac{X_{n-j+1}}{j}\right]\right] \\
&= \frac{n}{(n+4)} \frac{(n-1)}{(n+3)} \cdots \frac{(m+j)}{(m+j+4)} \left[\frac{(m+j-1)}{(m+j+1)} \cdots \frac{m}{(m+2)}\right. \\
&\quad + \frac{(m+j-1)}{(m+j+3)} \cdots \frac{m}{(m+4)} \frac{(m-1)}{(m+1)} \cdots \frac{j}{(j+2)} \\
&\quad \left. - 2\frac{(m+j-1)}{(m+j+2)} \cdots \frac{m}{(m+3)} \frac{(m-1)}{m} \cdots \frac{j}{(j+1)}\right] \\
&= \frac{m(m+1)(m+2)(m+3) - (4m^2 + 12m + 6)j + 2j^2}{(n+1)(n+2)(n+3)(n+4)}.
\end{aligned}$$

Using the identity $\mathrm{Cov}[x,y] = E[xy] - E[x]E[y]$ in (17) and replacing the expectations by the expressions just derived, one obtains (16) after lengthy manipulations (we used *Mathematica* for this).

Formula (15) can be obtained in a similar way, using the variance development:

$$\begin{aligned}
\frac{1}{n^4} \mathrm{Var}[S_{m,n}] &= \mathrm{Var}\left[\sum_{i=0}^{n-m+1} G_{m,i}^2\right] \\
&= (n-m+2)\ \mathrm{Var}[G_{m,0}^2] + 2\sum_{j=1}^{m-1} (n-m-j+2)\ \mathrm{Cov}[G_{m,0}^2 G_{m,j}^2] \\
&\quad + [(n-m+2)(n-m+1) + (m-1)(3m-2n-4)]\ \mathrm{Cov}[G_{m,0}^2 G_{m,m}^2]. \ \square
\end{aligned}$$

The asymptotic mean and variance constants $\mu_m$ and $\sigma_m^2$ of Proposition 1 are as follows. For $h(x) = \ln(x)$, one has $\mu_m = R(1, m-1) - \gamma$ and $\sigma_m^2 = (2m(m-1)+1)(\pi^2/6 - Q(1, m-1)] - 2m + 1$ (see Cressie 1976 and Holst 1979). For $h(x) = x^2$, one has $\mu_m = m(m+1)$ and $\sigma_m^2 = (2/3)m(m+1)(2m+1)$ (see also Cressie 1979).

Some authors (e.g., Dudewicz and van der Meulen 1981, Gokhale 1983) consider the following goodness-of-fit statistic:

$$L^{(e)}_{m,n} = \sqrt{6m/n}\left[\sum_{i=1}^{n} \ln(n\tilde{G}_{2m,i}) + n(\gamma - R(1, 2m-1))\right] \tag{18}$$

where

$$\tilde{G}_{2m,i} = \begin{cases} U_{(i+m)} - U_{(1)} & \text{for } i \le m; \\ U_{(i+m)} - U_{(i-m)} & \text{for } m < i < n; \\ U_{(n)} - U_{(i-m)} & \text{for } i \ge n-m. \end{cases} \tag{19}$$

Dudewicz et al. (1995) use the statistic (18) to test random number generators. This statistic is a normalization of an entropy estimator, defined by the inside sum divided by $n$, minus $\ln(2m)$. Under $H_0$, as $n, m \to \infty$ and $m/n \to 0$, this entropy estimator converges in probability to zero, which is the (differential) entropy of the uniform distribution (see Vasicek 1976). Dudewicz and van der Meulen (1981) have shown that $L^{(e)}_{m,n} \Rightarrow N(0,1)$ when $n \to \infty$ and $n^{\delta-1/3}m \to 0$ for some $\delta > 0$. Except for edge effects due to a different definition of the spacings for $i$ near 1 and $n$, $L^{(e)}_{m,n}$ is equivalent, up to a linear transformation, to the sum of logarithms of spacings. These edge effects are important only when $m/n$ decreases very slowly.

## 3 How good is the normal approximation ?

In this section, we want to assess how well the standard normal approximates the true distribution of the standardized statistics for finite $n$. To measure the error, we use the *Kolmogorov distance* between the two corresponding distribution functions. For any normalized statistic $Z$ (e.g., $Z = L^{(A)}_{m,n}$, or $Z = L^{(A,c)}_{m,n}$, and so on), define the error bounds:

$$\begin{aligned} \Delta^+(Z) &\stackrel{\text{def}}{=} \sup_{0\le u\le 1} (P[\Phi(Z) \le u] - u), \\ \Delta^-(Z) &\stackrel{\text{def}}{=} \sup_{0\le u\le 1} (u - P[\Phi(Z) \le u]), \end{aligned}$$

where $\Phi$ denotes the standard normal distribution.

These bounds are unknown but they can be estimated by Monte Carlo simulation, using Kolmogorov-Smirnov (KS) statistics based on the empirical distribution function $\hat{F}_N$ of $N$ i.i.d. copies of $\Phi(Z)$:

$$\begin{aligned} \hat{\Delta}^+(Z) &\stackrel{\text{def}}{=} \sup_{0\le u\le 1} (\hat{F}_N(u) - u), \\ \hat{\Delta}^-(Z) &\stackrel{\text{def}}{=} \sup_{0\le u\le 1} (u - \hat{F}_N(u)). \end{aligned}$$

We did this with $N = 10^6$, $n = 100, 1000, 10000$, and $m = 1, 4, 10, 20, 30$. With this value of $N$, the half-width of a 95% confidence interval on a given $\Delta^+(Z)$ or $\Delta^-(Z)$ is approximately 0.001. We thus give three fractional digits in Table 1. This table gives the error estimates for the four circular statistics, as well as for the entropy-based statistic $L^{(e)}_{m,n}$ defined in (18). For the latter, we used only $N = 10^5$ and give only two (significant) fractional digits, which is enough to observe large approximation errors. In this case, the half-width of a 95% confidence interval is 0.004.

For these Monte Carlo simulations, we used the combined multiple recursive generator proposed in L'Ecuyer (1996a). We also tried other robust generators of different natures, such as those given by L'Ecuyer (1988) and L'Ecuyer (1996b), and the results agreed up to the precision mentioned above.

We see from this table that the errors are significantly larger when using the asymptotic means and variances (the $A$ version) than when using their exact counterparts (the $E$ version). The distribution of $\Phi(L^{(A,c)}_{m,n})$ is *below* the uniform distribution, which means that $L^{(A,c)}_{m,n}$ is stochastically larger than a $N(0,1)$ (its distribution is more to the right). For $S^{(A,c)}_{m,n}$ and $L^{(e)}_{m,n}$, the opposite is true. These results indicate that the exact means and variances converge rather slowly to their asymptotic values, especially for large $m$. This can also be seen by looking at the exact and asymptotic formulæ. The values for the linear statistics (not reported here) are approximately the same as for the circular ones. For $L^{(e)}_{m,n}$, the approximation is very bad, due to bad centering and scaling. In other experiments, we found that replacing the $\tilde{G}_{2m,i}$ in (18) by their *circular* versions yields a significant improvement. However, $L^{(E,c)}_{m,n}$ is much better and should be used instead.

## 4 Discretization Effects

When the uniform random numbers are replaced by their finite-precision representations, as is the case for floating-point numbers on computers, then the small spacings can be exactly zero. Under $\mathcal{H}_0$, this happens with nonzero probability and when it happens, $L_{m,n}$, $L^{(c)}_{m,n}$, and $L^{(e)}_{m,n}$ become infinite; i.e., these statistics have an unbounded expectation. So, discretization has a very damaging effect. To prevent this, a simple heuristic is to reset any such null spacing to some small positive value $\epsilon$, smaller than the distance between the discrete output values

Table 1: Error bound estimates ($\pm 0.004$ for $L^{(e)}_{m,n}$, $\pm 0.001$ for others)

| | $n = 100$ | | $n = 1000$ | | $n = 10000$ | |
|---|---|---|---|---|---|---|
| $m$ | $\hat{\Delta}^+(L^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(E,c)}_{m,n})$ | $\hat{\Delta}^+(L^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(E,c)}_{m,n})$ | $\hat{\Delta}^+(L^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(E,c)}_{m,n})$ |
| 1 | 0.009 | 0.018 | 0.003 | 0.006 | 0.001 | 0.002 |
| 4 | 0.021 | 0.037 | 0.005 | 0.013 | 0.002 | 0.005 |
| 10 | 0.040 | 0.061 | 0.010 | 0.019 | 0.003 | 0.006 |
| 20 | 0.072 | 0.088 | 0.014 | 0.027 | 0.004 | 0.008 |
| 30 | 0.102 | 0.112 | 0.019 | 0.033 | 0.005 | 0.011 |
| $m$ | $\hat{\Delta}^+(L^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(A,c)}_{m,n})$ | $\hat{\Delta}^+(L^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(A,c)}_{m,n})$ | $\hat{\Delta}^+(L^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(L^{(A,c)}_{m,n})$ |
| 1 | 0.005 | 0.043 | 0.001 | 0.014 | 0.001 | 0.005 |
| 4 | 0.011 | 0.102 | 0.002 | 0.034 | 0.001 | 0.011 |
| 10 | 0.024 | 0.169 | 0.004 | 0.053 | 0.001 | 0.017 |
| 20 | 0.048 | 0.247 | 0.007 | 0.075 | 0.002 | 0.024 |
| 30 | 0.074 | 0.310 | 0.010 | 0.093 | 0.002 | 0.030 |
| $m$ | $\hat{\Delta}^+(S^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(E,c)}_{m,n})$ | $\hat{\Delta}^+(S^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(E,c)}_{m,n})$ | $\hat{\Delta}^+(S^{(E,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(E,c)}_{m,n})$ |
| 1 | 0.054 | 0.033 | 0.021 | 0.010 | 0.007 | 0.003 |
| 4 | 0.056 | 0.034 | 0.022 | 0.010 | 0.006 | 0.003 |
| 10 | 0.069 | 0.047 | 0.025 | 0.013 | 0.008 | 0.004 |
| 20 | 0.091 | 0.074 | 0.030 | 0.016 | 0.010 | 0.005 |
| 30 | 0.112 | 0.101 | 0.036 | 0.020 | 0.012 | 0.005 |
| $m$ | $\hat{\Delta}^+(S^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(A,c)}_{m,n})$ | $\hat{\Delta}^+(S^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(A,c)}_{m,n})$ | $\hat{\Delta}^+(S^{(A,c)}_{m,n})$ | $\hat{\Delta}^-(S^{(A,c)}_{m,n})$ |
| 1 | 0.175 | 0.013 | 0.058 | 0.004 | 0.019 | 0.001 |
| 4 | 0.273 | 0.006 | 0.091 | 0.002 | 0.029 | 0.001 |
| 10 | 0.396 | 0.006 | 0.131 | 0.001 | 0.042 | 0.000 |
| 20 | 0.531 | 0.008 | 0.178 | 0.002 | 0.057 | 0.000 |
| 30 | 0.632 | 0.009 | 0.215 | 0.002 | 0.069 | 0.000 |
| $m$ | | | $\hat{\Delta}^+(L^{(e)}_{m,n})$ | $\hat{\Delta}^-(L^{(e)}_{m,n})$ | $\hat{\Delta}^+(L^{(e)}_{m,n})$ | $\hat{\Delta}^-(L^{(e)}_{m,n})$ |
| 1 | | | 0.08 | 0.01 | 0.05 | 0.02 |
| 4 | | | 0.22 | 0.00 | 0.07 | 0.00 |
| 10 | | | 0.62 | 0.00 | 0.22 | 0.00 |
| 20 | | | 0.98 | 0.00 | 0.54 | 0.00 |

that can be produced by the generator. We did this in our implementation, with $\epsilon$ chosen as described in the next section.

How often are null spacings likely to happen? Consider a generator that produces a sequence of integers $Y_i \in \{0, \ldots, M-1\} = \mathbb{Z}_M$ and outputs $Y_i/M$ at step 1. There are thus $M$ possible output values. We generate a sample of size $n$ and compute the spacings. Assuming that the $Y_i$ are i.i.d. uniform over $\mathbb{Z}_M$, what is the expected number of zero spacings? This is the same as asking: If we throw $n$ balls randomly into $M$ boxes, what is the expected number of collisions (i.e., number of balls falling in a box already occupied)? For a given box, if $n \ll M$, the number $X$ of balls falling into it is approximately Poisson with parameter $\lambda = n/M$, so the expected number of collisions in this box is $E[X-1] + P[X = 0] \approx \lambda - 1 + e^{-\lambda} = \lambda^2/2 - \lambda^3/6 + O(\lambda^4)$. The expected total number of collisions is then $E[C] \approx M\lambda^2/2 = n^2/(2M)$. For example, for $M = 2^{31}$ (a typical case) and $n = 2^\ell$, the expected number of collisions is approximately 1, 16, 256, 4096 for $\ell = 16, 18, 20, 22$, respectively. In this case, if $n = 2^{20}$ (approximately one million) numbers are generated and if all the zero spacings are turned into an $\epsilon \ll 2^{-31}$, the value of the statistic will be much smaller than expected. On the other hand, if the generator has period length $M - 1$ and produces each $Y_i$ once and only once over its full cycle, then there will never be a collision as long as $n < M$. In this case, if $n^2/(2M)$ is large, the number of zero spacings will be much smaller than expected and the sum of logarithms of spacings will tend to be larger than expected. It could be more appropriate to use tests based on the spacings between *discrete* random variables, but we do not pursue this in the present paper.

## 5 Test results

In this section, we apply statistical tests based on the two statistics $L_{m,n}^{(E,c)}$ and $S_{m,n}^{(E,c)}$ to a few random number generators listed in Table 2. This list is far from exhaustive and our intention here is not to recommend any specific generator, but rather to investigate the usefulness of these tests to find weaknesses in certain classes of generators. The generators G1 to G13 have short period lengths from nowadays point of view, so they cannot be warmly recommended. However, many of them are still in widespread use. G14 to G18 have reasonable period lengths, althought G14 and G15 have well-known statistical defects (Compagner 1995, Matsumoto and Kurita 1996).

Table 2: List of selected generators.

| | |
|---|---|
| G1. | LCG with $M = 2^{31} - 1$ and $a = 742938285$. |
| G2. | LCG with $M = 2^{31} - 1$ and $a = 630360016$. |
| G3. | LCG with $M = 2^{31} - 1$ and $a = 16807$. |
| G4. | LCG with $M = 2^{32}$, $a = 69069$, and $c = 1$. |
| G5. | LCG with $M = 2^{31}$ and $a = 65539$. |
| G6. | LCG with $M = 2^{31}$ and $a = 452807053$. |
| G7. | LCG with $M = 2^{31}$, $a = 1103515245$, $c = 12345$. |
| G8. | Implicit inversive with $M = 2^{31} - 1$ and $a_1 = a_2 = 1$. |
| G9. | Explicit inversive with $M = 2^{31} - 1$ and $a = b = 1$. |
| G10. | Explicit inversive with $M = 2^{31} - 1$, $a = 1$, and $b = 12345$. |
| G11. | Implicit inversive with $M = 2^{32}$, $a = b = 1$, and $z_0 = 5$. |
| G12. | Explicit inversive with $M = 2^{32}$, $a = 6$, and $b = 1$. |
| G13. | Modified explicit inversive with $M = 2^{32}$, $a = 6$, and $b = 1$. |
| G14. | GFSR-521 in the Appendix of Ripley (1990). |
| G15. | GFSR proposed in Kirkpatrick and Stoll (1981). |
| G16. | Combined LCG in Fig. 3 of L'Ecuyer (1988). |
| G17. | Combined MRG in Fig. 1 of L'Ecuyer (1996a). |
| G18. | Combined Tausworthe generator in Fig. 1 of L'Ecuyer (1996b). |

G1 to G7 are linear congruential generators (LCGs), based on a recurrence of the form $x_i = (ax_{i-1} + c) \bmod M$, with output $u_i = x_i/M$ at step $i$. G1 is recommended in Fishman and Moore III (1986), G2 in Law and Kelton (1991), G3 and G4 are in many software packages (Bratley, Fox, and Schrage 1987, Ripley 1990), G5 is the infamous RANDU (Law and Kelton 1991), G6 is the URN12 generator of Dudewicz et al. (1995), and G7 is the `rand` function in the standard library of the C language (Plauger 1992). The next six are inversive generators modulo $M$, whose output at step $i$ is $u_i = z_i/M$. G8 is an *implicit* inversive generator of the form $z_i = (a_1 + a_2 z_{i-1}^{-1}) \bmod M$, where $0^{-1} \bmod M$ is defined as 0 (see Eichenauer-Herrmann 1992). G9 and G10 are *explicit* inversive generators of the form $x_i = (ai + b) \bmod M$, $z_i = x_i^{-1} \bmod M = x_i^{M-2} \bmod M$ (Eichenauer-Herrmann 1992, Hellekalek 1995). G11 is an implicit inversive generator with modulus $M = 2^e$ and recurrence $z_i = T(z_{i-1})$, where $T(2^\ell z) = (a_1 + 2^\ell a_2 z^{-1}) \bmod 2^e$ for odd $z$ (see Eichenauer-Herrmann and Grothe 1992). G12 and G13 are explicit inversive generators with power-of-two modulus, defined in Eichenauer-Herrmann and Ickstadt (1994) and Eichenauer-Herrmann (1996), respectively. G14 is the GFSR generator based on the recurrence $x_i := x_{i-521} \oplus x_{i-32}$, where $\oplus$ denotes the bitwise exclusive-or, and with the initialization procedure given in the Appendix of Ripley (1990). G15 is another GFSR generator, given in Kirkpatrick and Stoll (1981). G16 and G17 are the combined LCG of L'Ecuyer (1988) and the combined MRG given in Figure 1 of L'Ecuyer (1996a). G18 is the combined Tausworthe generator in Figure 1 of L'Ecuyer (1996b).

We use a *two-level* goodness-of-fit testing procedure, as suggested in Knuth (1981) and L'Ecuyer (1992): Generate $N$ independent replicates of the statistic of interest, $L_{m,n}^{(E,c)}$ or $S_{m,n}^{(E,c)}$, and compare their empirical distribution to the $N(0,1)$ by a standard KS test, as follows. Let $T_{(1)} \le \ldots \le T_{(N)}$ be the $N$ ordered values of the normalized statistic and define

$$\begin{aligned} D_N^+ &= \max_{1\le j\le N}(j/N - \Phi(T_{(j)})), \\ D_N^- &= \max_{1\le j\le N}(\Phi(T_{(j)}) - (j-1)/N). \end{aligned}$$

The distribution of $D_N^+$ and $D_N^-$ under $H_0$ is given, e.g., in Durbin (1973). One rejects $H_0$ if the *significance level* (the probability under $H_0$ that one of these KS statistics takes a value larger than the one observed) is extremely close to 0 or 1. In case of doubt, the test can be replicated (independently) and $H_0$ is rejected if the significance level is consistently too close to 0 or 1.

For a given total sample size $nN$, one can take either a small number of replications with a large $n$, or a larger $N$ with a smaller $n$. Some generators have good "average" behavior over their entire period or over very large segments of their cycles, but bad behavior over shorter segments. A larger $N$ may then be better. On the other hand, the lack of normality under $H_0$ may invalidate the test if $n$ is too small and $N$ too large. To detect problems over larger segments, one may as well take $N = 1$ with a large $n$. In this case, one has $D_N^- = \Phi(T_{(1)}) = 1 - D_N^+$, and the significance levels $\delta^+$ and $\delta^-$ of the KS statistics are the same as those of $T_{(1)}$ (on the two sides). However, recall the discretization problem that affect $L_{m,n}^{(E,c)}$ for large $n$. Even if the expected number of zero spacings (approximately $n^2/(2M)$) is not too large, this problem may also resurface if $N$ is large and $\epsilon \ll 1/M$. For our tests, for the generators whose output values are multiples of $1/M$, we took $\epsilon = 1/(2M)$.

To each of the 18 generators in Table 2, we applied the tests based on $L_{m,n}^{(E,c)}$ and $S_{m,n}^{(E,c)}$, with $(N,n) = (1,10^6),\ (100,10^4),\ (100,10^5),\ (1000,10^5)$ and $m = 1, 10, 20, 30$. For all the generators that require one or more integer seed(s) in addition to the parameters mentioned in the table, we used the same seed: 12345 everywhere, at the beginning of the first test. The tests were applied simultaneously for all values of $m$. For each of the two tests and each combination of parameters $(N, n, m)$, we obtained the significance levels $\delta^+$ and $\delta^-$ of the KS statistics. Tables 3 and 4 report the significance levels smaller than 0.01 or larger than 0.99 (that is, the suspect values), for $m = 1$, $(N, n) = (1, 10^6)$, and $(N, n) = (1000, 10^5)$. The entries with non-suspect values are left blank in the tables and the generators with no suspect significance levels (in the sense just defined) do not appear in the tables.

Table 3: Test results for $m = 1$, $N = 1$, $n = 10^6$.

| | $L_{m,n}^{(E,c)}$ | | $S_{m,n}^{(E,c)}$ | |
|---|---|---|---|---|
| Generator | $\delta^+$ | $\delta^-$ | $\delta^+$ | $\delta^-$ |
| G2 | 0.9940 | 6.0E-3 | | |
| G5 | 1-1.6E-13 | 1.6E-13 | | |
| G6 | >1-E-15 | <E-15 | | |
| G7 | 0.9916 | 8.4E-3 | | |
| G8 | 0.9993 | 7.2E-4 | | |
| G9 | 0.9913 | 8.7E-3 | | |
| G10 | 0.9966 | 3.4E-3 | | |

Table 4: Test results for $m = 1$, $N = 10^3$, $n = 10^5$.

| | $L_{m,n}^{(E,c)}$ | | $S_{m,n}^{(E,c)}$ | |
|---|---|---|---|---|
| Generator | $\delta^+$ | $\delta^-$ | $\delta^+$ | $\delta^-$ |
| G1 | | 2.1E-4 | | |
| G2 | | 5.0E-3 | | |
| G3 | | 8.8E-4 | | |
| G5 | 1-2.5E-5 | <E-15 | | |
| G6 | | <E-15 | | |
| G7 | | 1.9E-3 | | |
| G8 | 0.9997 | 4.3E-6 | | 0.9972 |
| G9 | 0.9778 | 4.2E-4 | | |
| G10 | | 1.1E-5 | | |
| G11 | 4.8E-14 | 0.9944 | | |
| G12 | 0.9984 | 3.7E-3 | | |
| G13 | 0.9902 | | | |

One can see from the two tables that several generators fail the tests based on the sum of logarithms of spacings, for $m = 1$, while much fewer fail the tests based on the sum of squares of spacings. For $(N, n) = (100, 10^4)$, which gives the same total sample size as $(N, n) = (1, 10^6)$, there was only one suspect value with $m = 1$, namely $\delta^+ = 0.0041$ for $S_{m,n}^{(E,c)}$ applied to G2. For $(N, n) = (100, 10^5)$, there were no suspect value for $S_{m,n}^{(E,c)}$ and 4 for $L_{m,n}^{(E,c)}$, for G3, G5, G6, and G11.

The results for $m > 1$ are not reported in the table, but they can be summarized as follows. Generally, there was much less suspect values than for $m = 1$, and the generators with suspect values for $m > 1$ had significance levels that were still more suspect for $m = 1$. So, $m = 1$ appears more sensitive, and the sum of logarithms of spacings appears more sensitive than the sum of squares of spacings to detect deficiencies in those linear congruential and inversive generators.

Generators G1 to G7 also fail other tests; see, e.g., L'Ecuyer (1992), L'Ecuyer, Cordeau, and Simard (1996), L'Ecuyer, Compagner, and Cordeau (1996), Leeb and Wegenkittl (1997). G5 and G6 are rather bad, so their failure is not surprising, but it is reassuring to see that the tests detect their weaknesses in such a spectacular way. The inversive generators G8 to G13 have difficulty with these spacings tests. In previous experiments by other authors (Hellekalek 1995, Leeb and Wegenkittl 1997), these inversive generators were much more successful than the LCGs. However, some of them also fail certain entropy tests introduced in L'Ecuyer, Compagner, and Cordeau (1996). The GFSR and combined generators G14 to G18 did well.

Before beeing recommended for general use, random number generators should successfully go through an extensive battery of statistical tests, with different types of tests to detect different types of defects. The sum of logarithms and sum of squares of spacings are logical candidates to enrich the existing collection of statistical tests for random number generators.

## Acknowledgments

This work has been supported by NSERC-Canada grant # ODGP0110050 and SMF0169893, FCAR-Québec grant # 93ER1654, Jean-François Cordeau, Yanick Champoux, and Richard Simard helped writing some of the computer programs and simplifying the variance expressions with *Mathematica*. I thank an anonymous referee for pointing out several references.

## REFERENCES

Bratley, P., B. L. Fox., and L. E. Schrage. 1987. *A Guide to Simulation.* Second ed. New York: Springer-Verlag.

COMPAGNER, A. 1995. Operational conditions for random number generation. *Physical Review E*, **52**(5-B), 5634–5645.

CRESSIE, N. 1976. On the logarithms of high-order spacings. *Biometrika*, **63**(2), 343–355.

CRESSIE, N. 1979. An optimal statistic based on higher-order gaps. *Biometrika*, **66**, 619–627.

DEL PINO, G. E. 1979. On the asymptotic distribution of $k$-spacings with applications to goodness-of-fits tests. *Annals of Statistics*, **7**, 1058–1065.

DUDEWICZ, E. J., AND E. C. VAN DER MEULEN. 1981. Entropy-based tests of uniformity. *Journal of the American Statistical Association*, **76**(376), 967–974.

DUDEWICZ, E. J., E. C. VAN DER MEULEN., M. G. SRIRAM., AND N. K. W. TEOH. 1995. Entropy-based random number evaluation. *American Journal of Mathematical and Management Sciences*, **15**, 115–153.

DURBIN, J. 1973. *Distribution Theory for Tests Based on the Sample Distribution Function*. volume 9 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.

EICHENAUER-HERRMANN, J. 1992. Inversive congruential pseudorandom numbers: A tutorial. *International Statistical Reviews*, **60**, 167–176.

EICHENAUER-HERRMANN, J. 1996. Modified explicit inversive congruential pseudorandom numbers with power-of-two modulus. *Statistics and Computing*, **6**, 31–36.

EICHENAUER-HERRMANN, J., AND H. GROTHE. 1992. A new inversive congruential pseudorandom number generator with power of two modulus. *ACM Transactions on Modeling and Computer Simulation*, **2**(1), 1–11.

EICHENAUER-HERRMANN, J., AND K. ICKSTADT. 1994. Explicit inversive congruential pseudorandom numbers with power of two modulus. *Mathematics of Computation*, **62**(206), 787–797.

FISHMAN, G. S., AND L. S. MOORE III. 1986. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal on Scientific and Statistical Computing*, **7**(1), 24–45.

GOKHALE, D. V. 1983. On entropy-based goodness-of-fit tests. *Computational Statistics and Data Analysis*, **1**, 157–165.

GREENWOOD, M. 1946. The statistical study of infectious diseases. *Journal of the Royal Statistical Society*, **109**, 85–103.

HALL, P. 1984. Limit theorems for sums of general functions of $m$-spacings. *Math. Proc. Cambridge Phil. Society*, **96**, 517–532.

HALL, P. 1986. On powerful distributional tests based on sample spacings. *Journal of Multivariate Analysis*, **19**, 201–224.

HELLEKALEK, P. 1995. Inversive pseudorandom number generators: Concepts, results, and links. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 255–262. IEEE Press.

HOLST, L. 1979. Asymptotic normality of sum-functions of spacings. *Annals of Probability*, **7**(6), 1066–1072.

JAMMALAMADAKA, S. R., X. ZHOU., AND R. C. TIWARI. 1989. Asymptotic efficiencies of spacing tests for goodness of fits. *Metrika*, **36**, 355–377.

KIRKPATRICK, S., AND E. STOLL. 1981. A very fast shift-register sequence random number generator. *Journal of Computational Physics*, **40**, 517–526.

KNUTH, D. E. 1981. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Second ed. Reading, Mass.: Addison-Wesley.

KUO, M., AND J. S. RAO. 1981. Limit theory and efficiencies for tests based on higher order spacings. In *"Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions*, 333–352. Calcuta: Statistical Publishing Society.

LAW, A. M., AND W. D. KELTON. 1991. *Simulation Modeling and Analysis*. Second ed. New York: McGraw-Hill.

L'ECUYER, P. 1988. Efficient and portable combined random number generators. *Communications of the ACM*, **31**(6), 742–749 and 774. See also the correspondence in the same journal, 32, 8 (1989) 1019–1024.

L'ECUYER, P. 1992. Testing random number generators. In *Proceedings of the 1992 Winter Simulation Conference*, 305–313. IEEE Press.

L'ECUYER, P. 1994. Uniform random number generation. *Annals of Operations Research*, **53**, 77–120.

L'ECUYER, P. 1996a. Combined multiple recursive generators. *Operations Research*, **44**(5), 816–822.

L'ECUYER, P. 1996b. Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, **65**(213), 203–213.

L'Ecuyer, P., A. Compagner., and J.-F. Cordeau. 1996. Entropy-based tests for random number generators. Submitted. Also GERAD technical report number G-96-41.

L'Ecuyer, P., J.-F. Cordeau., and R. Simard. 1996. Close-neighbor tests for random number generators. In preparation.

Leeb, H., and S. Wegenkittl. 1997. Inversive and linear congruential pseudorandom number generators in selected empirical tests. *ACM Transactions on Modeling and Computer Simulation*. To appear.

Marsaglia, G. 1985. A current view of random number generators. In *Computer Science and Statistics, Sixteenth Symposium on the Interface*, 3–10, North-Holland, Amsterdam. Elsevier Science Publishers.

Matsumoto, M., and Y. Kurita. 1996. Strong deviations from randomness in $m$-sequences based on trinomials. *ACM Transactions on Modeling and Computer Simulation*, **6**(2), 99–106.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods.* volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics.* Philadelphia: SIAM.

Plauger, P. J. 1992. *The Standard C Library.* Englewood Cliffs, New Jersey: Prentice Hall.

Pyke, R. 1965. Spacings. *Journal of the Royal Statistical Society, Series B*, **27**, 395–449.

Rao, J. S., and M. Kuo. 1984. Asymptotic results on the greenwood statistic and some of its generalizations. *Journal of the Royal Statistical Society, Series B*, **46**, 228–237.

Rényi, A. 1953. On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungaricae*, **4**, 191–231.

Ripley, B. D. 1990. Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics*, **31**, 153–163.

Russell, G. S., and D. J. Levitin. 1995. An expanded table of probability values for Rao's spacing test. *Communications in Statistics—Simulation and Computation*, **24**, 879–888.

Sethuraman, J., and J. S. Rao. 1970. Pitman efficiencies of tests based on spacings. In *Nonparametric Techniques in Statistical Inference*, ed. M. L. Puri, 405–515. Cambridge University Press.

Vasicek, O. 1976. A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B*, **38**, 54–59.