The Annals of Probability
1977, Vol. 5, No. 1, 72-81

THE DISTRIBUTION OF LEADING DIGITS AND UNIFORM DISTRIBUTION MOD 1[1]

BY PERSI DIACONIS

Stanford University

The lead digit behavior of a large class of arithmetic sequences is determined by using results from the theory of uniform distribution mod 1. Theory for triangular arrays is developed and applied to binomial coefficients. A conjecture of Benford's that the distribution of digits in all places tends to be nearly uniform is verified.

1. Introduction. A widely quoted empirical observation is that randomly occurring tables of data tend to have entries that begin with low numbers. There have been many theoretical models offered which predict that the proportion of entries beginning with first digit i is well approximated by

$$\log_{10}\left(1 + \frac{1}{i}\right), \qquad i = 1, 2, \cdots, 9 . \tag{1.1}$$

Excellent detailed surveys of the literature on this problem are in Knuth (1971) and Raimi (1976).

Almost the only large data set collected and referred to is the sample of 20,229 observations classified into 20 data types published by Benford (1938). Close to 30% of the data Benford presents comes from arithmetical sequences. The chi-squared statistic for goodness of fit of Benford's arithmetical sequences to the model (1.1) is greater than 440 on 8 degrees of freedom. This suggests such sequences as candidates for detailed mathematical analysis.

It is difficult to determine a complete list of the sequences Benford considered: $1/n$, $n^{\frac{1}{2}}$, a^n for a fixed, and $n!$ are listed, but undoubtedly others were also used. A standard set of tables, Abramowitz and Stegun (1964), yields the sequences in Table 1 below.

The theorems that follow yield the first digit behavior for each of the sequences in Table 1. The principal tool, Theorem 1, relates lead digit behavior to uniform distribution mod 1. Binomial coefficients $\binom{n}{k}$ for various values of n and $k = 0$ to n are given special treatment as a triangular array in Section 3 which provides a proof of a conjecture of Sarkar (1973). Benford (1938) conjectured that if all digits in a table are considered, not just lead digits, the relative frequency of the digits 0 through 9 approaches the uniform limit $\frac{1}{10}$ for large data sets. This is given a mathematical formulation and proof in Section 4. An argument due to

Received December 12, 1975; revised September 5, 1976.

[1] Research sponsored by National Science Foundation Grant MPS 7509450-000.

AMS 1970 *subject classifications.* Primary 10K05; Secondary 60F05.

Key words and phrases. Lead digits, uniform distribution mod 1, probabilistic number theory, Stei n's method for dependent variables.

TABLE 1
Numerical sequences from Section 1 *of Abramowitz and Stegun* (1964)

n^b	for b a fixed real number
a^n	for $a > 0$ fixed
$n\pi$	
$\log_b n$	for various bases b
$n!$	
p	p a prime
$\log p$	p a prime

Charles Stein leads to the variance and a central limit theorem for the number of ones in the binary expansion of a random integer.

2. Density and uniform distribution mod 1. Throughout, $[x]$ means the greatest integer less than or equal to x; $\langle x \rangle = x - [x]$ is the fractional part of x. In discussing the uniform distribution of sequences mod 1, the notation of Kuipers and Niederreiter (1974) will be followed. Let the left-most digit of a real number be defined by taking its left-most digit when expressed base 10 when sign and leading zeros are neglected (thus .008 has first digit 8). Digits k from the left are similarly defined. Without loss of generality, only positive numbers will be considered when discussing leading digit behavior.

The problem of a suitable definition of "pick an integer at random" has caused considerable difficulty at the foundational level. See Rényi (1970) page 73 and de Finetti (1972), pages 86, 98, 134. For consideration of numerical sequences, the naturally associated (finitely additive) measure is density or relative frequency. Let $D_1 = \{1, 2, \cdots, 9\}$, $D_i = \{0, 1, \cdots, 9\}$ for $i \geqq 2$, write $s_k = \prod_{i=1}^{k} D_i$. For $a > 0$, $x \in s_k$, let $a(\mathbf{x})$ be 1 if the jth digit from the left of a is x_j for $j = 1, \cdots, k$; let $a(\mathbf{x})$ be 0 otherwise.

Definition. A sequence of real numbers $\{a_i\}$ is a strong Benford sequence if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i(\mathbf{x}) = l(\mathbf{x})$$

for each $\mathbf{x} \in s_k$ and all integers k where

$$l(\mathbf{x}) = \log_{10}\left(\sum_{i=1}^{k-1} \frac{x_i}{10^{i-1}} + \frac{x_k{}^{+1}}{10^{k-1}}\right) - \log_{10}\left(\sum_{i=1}^{k} \frac{x_i}{10^{i-1}}\right).$$

Taking $k = 1$ in the definition shows that for a strong Benford sequence the relative frequency of lead digits i approaches $\log_{10}(1 + 1/i)$.

Several writers who discuss lead digit behavior have used the tools of uniform distribution mod 1 (see, for example, Feller (1971), page 63, or Macon and Moser (1950)). Raimi (1976), in a discussion of this point, mentions that strong Benford sequences were first defined by J. Cigler using the language of uniform distribution. The equivalence of the two definitions is given in Theorem 1.

THEOREM 1. *The sequence* $\{a_i\}$ *is a strong Benford sequence if and only if* $\{\log_{10} a_i\}$ *is uniformly distributed* mod 1.

PROOF. For fixed k let $\mathbf{x} \in s_k$. $a_i(\mathbf{x}) = 1$ if and only if $x_1 10^j + x_2 10^{j-1} + \cdots + x_k 10^{j-(k-1)} \leqq a_i < x_1 10^j + \cdots + (x_{k+1}) 10^{j-(k-1)}$ for some integer j. This holds if and only if

$$\log_{10}\left(x_1 + \cdots + \frac{x_k}{10^{k-1}}\right) \leqq \langle \log_{10} a_i \rangle < \log_{10}\left(x_1 + \cdots + \frac{x_{k+1}}{10^{k-1}}\right).$$

If $\log_{10} a_i$ is uniformly distributed mod 1, clearly

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i(\mathbf{x}) = l(\mathbf{x}).$$

Conversely, any interval $[a, b)$, $0 \leqq a < b \leqq 1$, can be approximated arbitrarily closely by a finite union of intervals with end points of the form

$$\log_{10}\left(x_1 + \frac{x_2}{10} + \cdots + \frac{x_k}{10^{k-1}}\right).$$

If a_i is a strong Benford sequence, the proportion of points of the sequence $\{\log_{10} a_i\}$ falling into intervals with such logarithmic end points will be proportional to the interval's length. Thus the sequence $\{\log_{10} a_i\}$ must be uniformly distributed mod 1. $\square$

Using known results from the theory of uniform distribution mod 1, the lead digit behavior of the sequences of Table 1 can be derived. For example, Theorem 2.6 of Chapter 1 of Kuipers and Niederreiter (1974) yields the following corollary:

COROLLARY 2. *If a sequence of real numbers* $\{a_i\}$ *is a strong Benford sequence, then* $\limsup_{n\to\infty} n \log (a_{n+1}/a_n) = \infty$.

Corollary 2 easily yields the fact that the sequences $\{n^b\}$ for any fixed real b, $\{an\}$ for $a > 0$ and $\{\log_b n\}$ for any base b are not strong Benford sequences. Using bounds from the prime number theorem, it is straightforward to show that the sequences $\{P_i\}$ and $\{\log P_i\}$, where P_i denotes the ith prime, are not strong Benford sequences.

For a an irrational number it is well known that the sequence $\{an\}$ is uniformly distributed mod 1. Thus Theorem 1 proves that $\{2^n\}$ is a strong Benford sequence.

Benford used the sequence $n!$ in his collection of numerical sequences. The next theorem establishes a conjecture of Sarkar (1973), who analyzed the first digits of factorials from 1 to 10,000.

THEOREM 3. *The sequence* $n!$ *is a strong Benford sequence.*

PROOF. Using Theorem 1, it must be shown that $\log_{10} n!$ is uniformly distributed mod 1. Using Stirling's formula and Theorem 1.2 of Chapter 1 of Kuipers and Niederreiter (1974), it is sufficient to show $\{f(n)\}$ is uniformly distributed mod 1 where $f(x) = a(x + \frac{1}{2}) \cdot \log x + bx$ for a and b constants. Straightforward differentiation and an application of Van der Corput's basic estimate for trigonometric sums (see, for example, Theorem 2.7 of Chapter 1 in Kuipers and Niederreiter (1974)) yield for any integer h

$$\frac{1}{n} |\sum_{j=1}^{n} \exp(2\pi i h f(j))| = O\left(\frac{\log n}{n^{\frac{1}{2}}}\right)$$

as $n \to \infty$. This gives the result by Weyl's criterion, Theorem 2.1 of Chapter 1 in Kuipers and Niederreiter (1974). □

For sequences without the strong Benford property, the relative frequency of lead digits does not tend to a limit as more and more terms are considered. While some easily accessible sequences have the strong Benford property, Benford's (1938) finding (page 556) that "the greatest variations from the logarithmic relation were found in the first digits of mathematical tables from engineering handbooks" suggests that he chose sequences without the strong Benford property.

3. Triangular arrays. Sequences such as n^k or binomial coefficients suggest consideration of triangular arrays. Results of this section are used to show the triangular array of binomial coefficients has the strong Benford property.

Let $\{a_{ij}\}$; $i = 1, 2, 3, \cdots$; $j = 1, 2, \cdots, f(i)$ be a triangular array of positive real numbers. Let $\chi_{[a,b)}$ be the indicator function of the interval $[a, b)$, $0 \leqq a \leqq b \leqq 1$.

DEFINITION. The array $\{a_{ij}\}$ is uniformly distributed mod 1 if

$$\lim_{i\to\infty} \frac{1}{f(i)} \sum_{j=1}^{f(i)} \chi_{[a,b]}(\langle a_{ij}\rangle) = b - a$$

for each $0 \leqq a < b < 1$.

REMARKS.

1. The principal application of the definitions of this section will be triangular arrays with $f(k) = k$. Pólya and Szegö (1972), page 93, prove results which imply that the array of the Farey sequence and the array j/n, where j runs through the $\phi(n)$ numbers relatively prime to n ($\phi(n)$ is Euler's function), are uniformly distributed mod 1.

2. The distribution of left-most digits may be analogously defined for two-dimensional arrays and, with obvious modifications, a version of Theorem 1 above connects the distribution of leading digits and uniform distribution of triangular arrays. For this reason, only the uniform distribution versions of results will be given.

3. If a sequence is formed from a triangular array by considering the array rows one after another, then if the array was uniformly distributed mod 1 it can be shown that the sequence will be uniformly distributed mod 1. The array

$$\begin{aligned} a_i &= 1 \qquad i \text{ even}, \quad 1 \leqq j \leqq i \\ &= 0 \qquad i \text{ odd}, \quad 1 \leqq j \leqq i \end{aligned}$$

shows the converse of this statement fails even for bounded sequences. A similar comment holds for the distribution of leading digits.

Sarkar (1973) computed the lead digit behavior for the array $\binom{n}{k}$ for $0 \leqq k \leqq n$, $n = 1$ to 500, and on the basis of numerical results conjectured that the triangular array would have the strong Benford property. Theorem 4 provides a proof of this in the language of uniform distribution mod 1.

THEOREM 4. *The triangular array* $\log \binom{n}{i}$ *is uniformly distributed* mod 1.

PROOF. Consider for an integer h, the obvious inequality

$$(3.1) \qquad \left|\sum_{j=0}^{n} \exp\left(2\pi i h \log \binom{n}{i}\right)\right| \leqq 3 + 2\left|\sum_{1 \leq j < n/2} \exp(-2\pi i h(\log j! + \log (n-j)!)\right| .$$

To bound the right side of (3.1) use Stirling's formula and a triangular-array version of Theorem 1.2 of Chapter 1 of Kuipers and Niederreiter (1974) to see that it is enough to evaluate the sum

$$(3.2) \qquad \sum_{1 \leq j < n/2} \exp(2\pi i f(j)) .$$

In (3.2) $f(x) = -x \log x + (n-x) \log (n-x)$, $f'(x) = \log (n-x) - \log x$, $f''(x) = 1/(n-x) - 1/x > c/n$ for some constant c. Thus, the standard Van der Corput arguments referred to in Theorem 3 show the sum in (3.2) is $O(n^{\frac{1}{2}} \log n)$. Dividing by n and using a triangular array version of Weyl's criterion gives the result. □

4. Another law given in Benford's paper. Benford and other writers who consider leading digits have noted that digits k places from the left have curious limiting behavior. The results are similar to those for lead digits. For example, the relative frequency of integers with second digit equal to 1 does not tend to a limit as the number of integers considered goes to infinity. Any of the approaches used for lead digits can be employed to suggest the limit

$$\log_{10}\left(\frac{12 \cdot 22 \cdots 92}{11 \cdot 21 \cdots 91}\right) \doteq .1139 .$$

The suggested probabilities for second digit 0 to 9 are more nearly equal than the corresponding probabilities for lead digits. Looking further to the right of the lead digit, the suggested probabilities can be shown to get closer and closer to $\frac{1}{10}$. Benford (1938), page 553, writes: "As a result of this approach to the uniformity in the qth place, the distribution of digits in all places in an extensive tabulation of multi-digit numbers will be nearly uniform."

In this case, a theorem can be given which substantiates Benford's conjecture. Let $n = \sum_{k=0}^{\infty} \varepsilon_k(n) 10^k$ be the expansion of n to base 10. Thus $0 \leqq \varepsilon_k(n) \leqq 9$ and for fixed n only finitely many $\varepsilon_k(n)$ are different from 0. Let $Y(k)$ be the number of digits of the integer k in the base 10. Let $D(x) = \sum_{k \leq x} Y(k)$ be the number of all digits of integers less than or equal to x. For $0 \leqq k \leqq 9$ let $D_k(x)$ be the number of occurrences of the digit k in all numbers less than or equal to x.

THEOREM 5.

$$\frac{D_k(x)}{D(x)} = \frac{1}{10} + O\left(\frac{1}{\log x}\right) \qquad as \quad x \to \infty .$$

PROOF. For simplicity, the result will be proved when $k = 1$.

$$D(x) = \sum_{j \leq x} Y(j) = \sum_{j \leq x} ([\log_{10} j] + 1] = x \log_{10} x + O(x) . \tag{4.1}$$

Let $\varepsilon_k'(j)$ be 1 if the coefficient of 10^k in the base 10 expansion of j is 1; $\varepsilon_k'(j) = 0$ otherwise;

$$\begin{aligned} D_1(x) &= \sum_{j \leq x} \sum_{k=0}^{\infty} \varepsilon_k'(j) = \sum_{k=0}^{[\log_{10} x]} \sum_{j \leq x} \varepsilon_k'(j) \\ &= \sum_{k=0}^{[\log_{10} x]} \sum_{i=8\cdot 10^k+1}^{9\cdot 10^k} \left[\frac{x+i}{10^{k+1}}\right] = \frac{x}{10} \log_{10}(x) + O(x) . \end{aligned} \tag{4.2}$$

Standard manipulation of the right-hand sides of (4.1) and (4.2) gives the theorem. □

The argument used to prove Theorem 5 can be applied to prove similar results for subsets of the integers such as the square-free numbers. Charles Stein has kindly provided a probabilistic proof of Theorem 5 using techniques similar to those in Stein (1970). His proof generalizes to give the variance of the number of ones. For simplicity, the proof is given for binary expansions.

Let X be an integer chosen uniformly on $[0, n]$. Let $Y = Y(X)$ be the number of ones in the binary expansion of X. In Lemma 6 and in Theorems 7 and 8, numbers $x \leqq n$ are written with $m = [\log_2 n] + 1$ digits, all leading zeros being counted as possible digits. Let $Q = Q(x, n)$ be the number of zeros in the binary expansion of x which cannot be changed into ones without making the transformed number greater than n (thus $Q(10, 5) = 1$).

LEMMA 6.

$$E(Y) = \tfrac{1}{2}(m - E(Q)) , \tag{4.3}$$

$$\operatorname{Var}(Y) = \frac{m}{4}\left\{1 - \frac{(E(Q) + 2\operatorname{Cov}(Y, Q))}{m}\right\} . \tag{4.4}$$

PROOF. Let I be a random subscript distributed uniformly on $[1, m]$. For $0 \leqq x \leqq n$, let x_i be the ith digit of x for $i = 1$ to m. Let

$$\begin{aligned} Y' &= Y + 1 - 2X_I \qquad && \text{if} \quad X + (1 - 2X_I)2^{I-1} \leqq n \\ &= Y && \text{otherwise.} \end{aligned}$$

Thus Y' is the number of digits in the number X' which has the same digits as X except that the Ith digit has been changed from X_I to $1 - X_I$. Y and Y' are exchangeable random variables. For any two exchangeable variables, the following identity holds provided the expectations involved exist:

$$\begin{aligned} 0 &= E\{(Y' - Y)(f(Y) + f(Y'))\} \\ &= E\{2(Y' - Y)f(Y) + (Y' - Y)(f(Y') - f(Y))\} . \end{aligned} \tag{4.5}$$

It is straightforward to check that

$$E\{(Y' - Y) \mid X\} = \frac{-Y}{m} + \left(1 - \frac{Y}{m}\right) - \frac{Q}{m}. \tag{4.6}$$

Taking the expectation of both sides of (4.6) and using (4.5) with $f(Y) \equiv 1$ gives (4.3).

To prove (4.4), notice that

$$E\{(Y - Y')^2 \mid X\} = 1 - \frac{Q}{m}. \tag{4.7}$$

Take $f(Y) = Y$ in (4.5) and use (4.7) and (4.6), yielding

$$E\left(1 - \frac{Q}{m}\right) = 2E\left\{Y\left(\frac{2Y}{m} - 1 + \frac{Q}{m}\right)\right\} = \frac{4 \operatorname{Var} Y}{m} + 2 \operatorname{Cov}\left(Y, \frac{Q}{m}\right). \tag{4.8}$$

Solving in (4.8) for Var Y yields (4.4). □

THEOREM 7. *With notation as above, the following two approximations are valid*:

$$\sum_{i \leq x} Y(i) = \frac{n \log_2 n}{2} + O(n); \tag{4.9}$$

$$\sum_{i \leq n} \left(Y(i) - \frac{\log_2 n}{2}\right)^2 = \frac{n \log_2 n}{4} + O(n(\log n)^{\frac{1}{2}}). \tag{4.10}$$

PROOF. Consider the random variable Q of Lemma 6. It is not hard to see that $Q(x) \geq k$ implies x coincides with n in its left-most $k - 1$ digits. The number of $x \leq n$ which coincide with n in the left-most $k - 1$ digits is bounded by $2^{[\log_2 n]-(k-1)}$. Thus

$$P(Q > k) \leq \frac{1}{2^k}. \tag{4.11}$$

It follows that

$$E(Q) = \sum_{k=0}^{\infty} P(Q > k) = O(1).$$

This and Lemma 6 prove (4.9). For (4.10), consider the inequality Cov $(Y, Q) \leq$ (Var $Y)^{\frac{1}{2}}$(Var $Q)^{\frac{1}{2}}$. Using (4.11) to show Var $Q = O(1)$, Lemma 6 yields

$$\operatorname{Var}(Y) = \frac{\log_2 (n)}{4} + O((\operatorname{Var} Y)^{\frac{1}{2}}).$$

A straightforward argument leads from this to (4.10). □

A final result obtained jointly with Professor Stein is the following central limit theorem for the number of ones in the binary expansion of a random integer.

THEOREM 8. *Let X be an integer chosen uniformly on* $[0, n]$. *Let Y be the number* *of ones in the base 2 expansion of X. Write* $m = [\log_2(n)] + 1$. *Then*

$$\sup_{-\infty<x<\infty}\left|P\left(\frac{Y - m/2}{(m/4)^{\frac{1}{2}}} \leqq x\right) - \Phi(x)\right| \leqq \frac{c}{m^{\frac{1}{2}}}$$

where $\Phi(x) = 1/(2\pi)^{\frac{1}{2}} \int_{-\infty}^{x} e^{-x^2/2}\,dx$ *and c is a constant.*

PROOF. With notation as in Lemma 6 and Theorem 7 above, let $S = (Y - m/2)/(m/4)^{\frac{1}{2}}$ and $S' = (Y' - m/2)/(m/4)^{\frac{1}{2}}$. Let the real-valued continuous function $g(x)$ be defined as

$$\begin{aligned} g(x) &= (2\pi)^{\frac{1}{2}}e^{x^2/2}\Phi(x)[1 - \Phi(b)] && \text{if } x \leqq b \\ &= (2\pi)^{\frac{1}{2}}e^{x^2/2}\Phi(b)[1 - \Phi(x)] && \text{if } x \geqq b\,. \end{aligned}$$

Stein (1970), page 594, has shown that g satisfies $g'(x) = xg(x) + \chi_b(x) - \Phi(b)$ where χ_b is the indicator function of the set $[-\infty, b]$. The differential equation is understood in the sense that the integral of the left-hand side between any two limits is equal to the integral of the right-hand side between the same limits. Stein (1970) has also proved that

$$(4.12) \qquad |g(x)| \leqq 1\,, \qquad |xg(x)| \leqq 1\,, \qquad |g'(x)| \leqq 1\,.$$

Using (4.5) yields, with $f(Y) = g(S)m^{\frac{1}{2}}/2$,

$$(4.13) \qquad 0 = E\left[\{m^{\frac{1}{2}}g(S)E(Y - Y' \mid X)\} - \frac{m^{\frac{1}{2}}}{2}(Y - Y')\{g(S) - g(S')\}\right].$$

Now (4.6) implies

$$m^{\frac{1}{2}}E[Y - Y' \mid X] = S + \frac{Q}{m^{\frac{1}{2}}}\,.$$

Using this in (4.13) along with $E(Q) = O(1)$ yields

$$(4.14) \qquad 0 = E\left\{Sg(S) - \frac{m^{\frac{1}{2}}}{2}(Y - Y')[g(S) - g(S')]\right\} + O\left(\frac{1}{m^{\frac{1}{2}}}\right).$$

Define the random step function $h(t)$ as

$$\begin{aligned} h(t) &= \frac{Y}{2m^{\frac{1}{2}}} && \text{if } S - \frac{2}{m^{\frac{1}{2}}} < t \leqq S \\ &= \frac{m - Y}{2m^{\frac{1}{2}}} && \text{if } S < t \leqq S + \frac{2}{m^{\frac{1}{2}}} \\ &= 0 && \text{elsewhere.} \end{aligned}$$

Note that $\int_{-\infty}^{\infty} h(t) = 1$ and that conditional on X

$$\frac{m^{\frac{1}{2}}}{2}(Y - Y') = 0 \quad \text{and} \quad S' = S \quad \text{with probability } \frac{Q}{m}$$

$$\frac{m^{\frac{1}{2}}}{2}(Y - Y') = \frac{m^{\frac{1}{2}}}{2} \quad \text{and} \quad S' = S - \frac{2}{m^{\frac{1}{2}}} \quad \text{with probability } \frac{Y}{m}$$

$$\frac{m^{\frac{1}{2}}}{2}(Y - Y') = \frac{-m^{\frac{1}{2}}}{2} \quad \text{and} \quad S' = S + \frac{2}{m^{\frac{1}{2}}} \quad \text{with probability } \frac{m - Y - Q}{m}\,.$$

Thus

$$E\left\{\frac{m^{\frac{1}{2}}}{2}(Y-Y')[g(S)-g(S')]\right\}$$

$$= E\left\{\int_{-\infty}^{\infty} g'(t)h(t)\,dt - \frac{Q}{2m^{\frac{1}{2}}}\left[g\left(S+\frac{2}{m^{\frac{1}{2}}}\right)-g(S)\right]\right\} \tag{4.15}$$

$$= E\{\textstyle\int_{-\infty}^{\infty}(tg(t)+\chi_b(t)-\Phi(b))h(t)\,dt\} + O\left(\frac{1}{m^{\frac{1}{2}}}\right).$$

Using (4.15) in (4.14) yields

$$0 = E\{Sg(S) - \textstyle\int_{-\infty}^{\infty}(tg(t)+\chi_b(t))h(t)\,dt\} + \Phi(b) + O\left(\frac{1}{m^{\frac{1}{2}}}\right). \tag{4.16}$$

It is straightforward to check that

$$P\left(S < b - \frac{2}{m^{\frac{1}{2}}}\right) \leqq E\{\textstyle\int_{-\infty}^{\infty}\chi_b(t)h(t)\,dt\} \leqq P\left(S < b + \frac{2}{m^{\frac{1}{2}}}\right). \tag{4.17}$$

Assume for the moment

$$E\{Sg(S) - \textstyle\int tg(t)h(t)\,dt\} = O\left(\frac{1}{m^{\frac{1}{2}}}\right). \tag{4.18}$$

Then (4.16) and (4.18) hold uniformly in b and imply there are positive constants e_1 and e_2 such that

$$-\frac{e_1}{m^{\frac{1}{2}}} \leqq E\{\textstyle\int_{-\infty}^{\infty}\chi_b(t)h(t)\,dt\} - \Phi(b) \leqq \frac{e_2}{m^{\frac{1}{2}}}.$$

From (4.17),

$$P(S<b) \leqq E\{\textstyle\int_{-\infty}^{\infty}\chi_{b+2/m^{\frac{1}{2}}}(t)h(t)\,dt\} \leqq \Phi\left(b+\frac{2}{m^{\frac{1}{2}}}\right)+\frac{e_2}{m^{\frac{1}{2}}}$$

$$\leqq \Phi(b) + \frac{e_3}{m^{\frac{1}{2}}}.$$

In the last inequality the easily verified fact that

$$\Phi(c) - \Phi\left(c \pm \frac{2}{m^{\frac{1}{2}}}\right) = O\left(\frac{1}{m^{\frac{1}{2}}}\right)$$

uniformly in c is used. Corresponding inequalities give the lower bound for $P(S<b)$ and the theorem. Thus it only remains to prove (4.18). Note that (4.12) implies $|g(x)-g(y)| \leqq |x-y|$ and thus $|sg(s)-tg(t)| \leqq (|s|+1)|s-t|$. This last bound implies

$$|E\{sg(s) - \textstyle\int tg(t)h(t)\,dt\}| \leqq E\{(|s|+1)\textstyle\int_{-\infty}^{\infty}|s-t|h(t)\,dt\}$$

$$= \frac{1}{m^{\frac{1}{2}}}E\{|s|+1\} = O\left(\frac{1}{m^{\frac{1}{2}}}\right)$$

since Theorem 9 implies $E\{|s|+1\} = O(E(s^2)^{\frac{1}{2}}) = O(1)$. □

Reference to number theoretic functions related to $D_1(x)$ may be found in Delange (1975).

Acknowledgments. Thanks are due to Professor Charles Stein for allowing publication of his results given as Lemma 6 and Theorem 7 and for his help in proving Theorem 8. Professor Patrick Billingsley's many detailed comments have been a great help.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* National Bureau of Standards.

[2] BENFORD, FRANK (1938). The law of anomalous numbers. *Proc. Amer. Philos. Soc.* **78** 551–572.

[3] DE FINETTI, B. (1972). *Probability Induction and Statistics.* Wiley, New York.

[4] DELANGE, H. (1975). Sur La Fonction Sommatoire De La Fonction "Somme Des Chiffres," L'Enseignement Mathematique II Serie, **21**, 31–47.

[5] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

[6] KNUTH, D. (1971). *The Art of Computer Programming* **2**. Addison-Wesley, Reading, Mass.

[7] KUIPERS, L. and NIEDERREITER, H. (1974). *Uniform Distribution of Sequences.* Wiley, New York.

[8] MACON, N. and MOSER, L. (1950). On the distribution of first digits of powers. *Scripta Math.* **16** 290–291.

[9] PÓLYA, G. and SZEGÖ, G. (1972). *Problems and Theorems in Analysis* **1**. Springer, Berlin.

[10] RAIMI, R. (1976). The first digit problem. *Amer. Math. Monthly* **83** 521–538.

[11] RÉNYI, A. (1970). *Probability Theory.* North-Holland, Amsterdam.

[12] SARKAR, P. B. (1973). An observation on the significant digits of binomial coefficients and factorials. *Sankhyā B* **35** 363–364.

[13] STEIN, C. (1970). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 583–602, Univ. of California Press.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305