International Journal of Pure and Applied Mathematics
Volume 7 No. 4 2003, 497-506

ON THE DISTRIBUTION OF log(p)

H.L. Hutson
Department of Mathematics
The US Merchant Marine Academy
Kings Point
Great Neck, NY 11024, USA
e-mail: hutsonh@usmma.edu

Abstract: A classical theorem of Franel says that the proportions of a given digit in the jth place after the decimal point in a table of log(n), n=1,2,3,... are dense in a non trivial interval. We revive this result by considering the subtleties that occur when one restricts attention to a table of log(p), p prime.

AMS Subject Classification: 11A41, 11K06
Key Words: prime numbers, uniform distribution modulo 1

1. Introduction

In a book entitled *La Science et l'Hypothèse*, Poincaré asked (see [3] and [10]) in regard to a table of logarithms "What is the probability that its third decimal is an even number?" and added "you would not hesitate to reply $\frac{1}{2}$". In fact this obvious answer is false in a very strong way. If you let $v_g^j(n)$ denote the number of $g = 0,1,2,3\ldots 9$, which appear in the j-th place after the decimal point among the first n entries in a table of log(n) then the quotients $\dfrac{v_g^j(n)}{n}$, $n = 1,2,3,\ldots$ are dense in a non trivial interval (see [11]). A simple argument shows that the same result holds for the proportions of entries which show an even digit.

Received: April 27, 2003

Problems like that of Poincaré are intimately related to the concept of uniform distribution modulo 1 (as general references on this subject we use [8] and [2]). Recall that a sequence x_n, $n = 1,2,3,\ldots$ is u.d. mod 1 if the proportion of the number $\{x_n\}$ (we denote the fractional part of x by $\{x\}$) among the first n entries of the sequence x_n which lie in $I = [a, b)$ is asymptoticly equal to b - a for all $[a, b) \subset [0, 1]$. In order to have notation for this let:

$$\phi(I, n, x_n) = \sum_{i=1}^{n} \chi_I(\{x_i\}) ,$$

where $\chi_I(x)$ denotes the characteristic function of I. We have that the following are equivalent:

1. x_n is u.d. mod 1.
2. $\lim\limits_{n\to\infty} \dfrac{\phi([a, b), n, x_n)}{n} = b - a \;\forall\; [a, b) \subset [0, 1]$.
3. $\lim\limits_{n\to\infty} \dfrac{\phi([0, a), n, x_n)}{n} = a \;\forall\; [0, a) \subset [0, 1]$.
4. For all R.I. functions $f(x)$ on $[0, 1]$,
$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} f(\{x_i\})}{n} = \int_0^1 f(x)\, dx .$$

The last equivalence is clear since (2) shows that (4) holds for step functions and any Riemann integrable function can be approximated by these.

The link of u.d. mod 1 with problems like that of Poincaré is now evident: Consider a table of values of the decimal expansions of $f(x)$. Let $x_n = 10^{j-1} f(n)$. Then if a_n^j denotes the digit in the jth place after the decimal point of the nth entry in the table we have $a_n^j = [10x_n]$ ($[x]$ denotes the integer part of x). Thus $a_g^j = g$ iff $x_n \in [\frac{g}{10}, \frac{g+1}{10})$ so that (now using the notation $v_g^j(n)$ to denote the number of g appearing among the first n entries in the j-th place after the decimal point of a table of $f(n)$)

$$\lim_{n\to\infty} \frac{v_g^j(n)}{n} = \lim_{n\to\infty} \frac{\phi([\frac{g}{10}, \frac{g+1}{10}), n, x_n)}{n} = \frac{1}{10} ,$$

whenever $10^{j-1} f(n)$ is u.d. mod 1. Note that the theorem stated above in answer to Poincaré question implies that $10^{j-1} \log(n)$ is not u.d. mod 1 for all $j \geq 1$.

Exponential sums, also, play an important role in the study of u.d. mod 1 [14]. We have that the following are equivalent (Weyl criterion):

1. x_n is u.d. mod 1.
2. $\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \exp(2\pi i h x_k)}{n} = 0$ for all integer $h \neq 0$.

This is clear from (4) of our previous equivalences since Riemann integrable functions can be approximated by trigonometric polynomials and since $\int_0^1 \exp(2\pi i h x)\, dx = 0$.

An exponential sum similar to that mentioned above was used by Turán to study the distribution of $\log(p)$. From the forties until his death in 1976 Turán enjoyed studying the distribution of the sequence $\log(p)$, p prime, with a view toward the Riemann Hypothesis (see [13], Chapters 33-35). His interest lay in his 1947 discovery of an interesting link between this sequence and the zero free regions of the Riemann zeta function. Many have hoped to clarify the relation between the primes and the zeros of the zeta function. In 1973 Turán said "we ask whether the zeta-roots of some ranges and the primes of some range influence each other particularly strongly, and if this is indeed the case then how?" Turán considered the sum

$$Z(\tau, N_1, N_2) = \sum_{N_1 \le p \le N_2} \exp(-i\tau \log(p))\,.$$

A remarkable discovery was the following:

Theorem 1.1. *Suppose the existance of constants $a \geq 2$ and $0 < b \leq 1$ and c(a,b) so that for all $\tau > c(a,b)$ the inequality*

$$|Z(\tau, N_1, N_2)| < \frac{N \log^{10}(N)}{\tau^b}$$

holds for all pairs (N_1, N_2) with

$$\tau^a \leq N \leq N_1 < N_2 \leq \exp(\tau^{b/10})\,,$$

then the half plane $x > 1 - e^{-10} b^3/a^2$ contains at most finitely many zeros of the Riemann zeta function.

Turán refers to such a result as a quasi-Riemann conjecture. Many related theorems and partial converses are also available.

Our interest in the sequence $\log(p)$ is more modest. First we review early work of Wintner [15] on the failure of uniform distribution modulo 1 for this sequence. Next we prove that the digits in a table of the decimal expansions of $f(p) = (\log(p))^{\beta}, \beta \in (0,1)$, have the property that

$$\frac{v_g^j(n)}{n}, \quad \text{n=1,2,3,...}$$

is dense in [0,1]. This illustrates the interesting behavior of the sequence at hand for as noted above uniform distribution modulo 1 of $10^{j-1}f(p)$ for $j \geq 1$ and some function $f(p)$ implies that $\frac{v_g^j(n)}{n} \to \frac{1}{10}$ (of course this is the distribution of the digits one would naively hope for) as n increases. Thus, indeed, u.d. mod 1 fails in a very strong way. In view of Weyl criterion this failure perhaps provides some insight into why Turán idea is both difficult and subtle in so far as the study of $Z(\tau, N_1, N_2)$ might be viewed as an attempt to understand the way in which u.d. mod 1 fails. We also provide a partial generalization of the result on the distribution of the digits stated above.

2. The Results

In 1935 Wintner [15] showed that the sequence $\log(p)$ is not u.d. mod 1 by computing its upper and lower asymptotic density functions. The only tool needed is the prime number theorem in its simplest form

$$\pi(x) \sim \frac{x}{\log(x)}\,,$$

where $\pi(x)$ denotes the number of primes $\leq x$. To understand Wintner result let $b \in [0,1]$ and

$$\phi(x,b) = \text{card}\{p|\ \log(p) \leq x,\ \{\log(p)\} \leq b\}.$$

Also, let

$$N(x) = \text{card}\{p|\ \log(p) \leq x\}.$$

Wintner computes

$$\phi_+(b) = \limsup_{x\to\infty} \frac{\phi(x,b)}{N(x)} \quad \text{and} \quad \phi_-(b) = \liminf_{x\to\infty} \frac{\phi(x,b)}{N(x)}$$

(these are called the upper an lower asymptotic density functions of the sequence). We have the following theorem.

Theorem 2.1. (Wintner 1935)

$$\phi_+(b) = \frac{1-e^{-b}}{1-e^{-1}} \quad \textit{and} \quad \phi_-(b) = \frac{e^b - 1}{e-1}.$$

In addition all ν *such that* $\Phi_-(b) \leq \nu \leq \Phi_+(b)$ *are limit points of* $\frac{\phi(x,b)}{N(x)}$ *as* x *passes through the real numbers.*

It is also true that u.d. mod 1 fails for $\sqrt{\log(p)}$ since uniformly distributed sequences have a minimum rate of growth. It is now natural to look at a general situation: let $v_g^j(n)$ denote the number of g, g=0,1,2,. . . 9, which appear amoung the first n entries in the j-th place after the decimal point in a table of decimal expansions of $f(n)$ for $n \in N$, where N is a subset of the natural numbers. If $f(x) = \sqrt{\log(x)}$ and we let N be the set of all natural numbers the quotients $\frac{v_g^j(n)}{n}$ are dense in [0,1] (see [11] page 92) and thus, u.d. mod 1 fails in a very strong way. With this in mind we provide the following generalization.

Theorem 2.2. [1] *Let $f(x) = (\log(x))^\beta$, where $\beta \in (0,1)$ and let N be the set of primes. The sequence $\frac{v_g^j(n)}{n}$ is dense in [0,1].*

Proof. Let $x_n = (\log(p))^\beta = a_0^n.a_1^n a_2^n \dots a_j^n \dots$ and $y_n = 10^{j-1}x_n - [10^{j-1}x_n]$. Then $[10y_n] = a_j^g = g$ holds iff $\frac{g}{10} \le y_n < \frac{g+1}{10}$. That is iff $\frac{g}{10} \le 10^{j-1}(\log(p_n))^\beta - k_n < \frac{g+1}{10}$, where $k_n = [10^{j-1}(\log(p_n)^\beta]$. Now, let $X_{k_n}(g,j) = \exp\left(\frac{g}{10^j} + k_n\right)^{\frac{1}{\beta}}$. Isolating p_n in the middle of the previous inequality we find that the digit in the j-th place after the decimal point of $(\log(x))^\beta$ is g iff $p_n \in [X_{k_n}(g,j), X_{k_n}(g+1,j))$. Note that $\frac{X_{k_n}(g,j)}{X_{k_n}(g+1,j)} = o(1)$ as $n \to \infty$. To see this expand the powers of e involved by the binomial theorem after factoring out $(k_n)^{\frac{1}{\beta}}$. Given $\alpha \in [0,1]$ we must find a sequence n_k with $\lim_{k\to\infty} \frac{v_g^j(n_k)}{n_k} = \alpha$. Observe that for large k_n we have $\frac{1}{\alpha}X_{k_n}(g+1,j) < X_{k_n}(g+2,j)$ since $\frac{X_{k_n}(g+2,j)}{X_{k_n}(g+1,j)} \to \infty$ as $n \to \infty$. Hence if $p_n \in [X_{k_n}(g+1,j), \frac{1}{\alpha}X_{k_n}(g+1,j))$ and k_n is large the digit in the j-th place after the decimal point in the expansion of $(\log(p_n))^\beta$ is g+1. Now let $n_k = n_{k_n} = \pi\left(\frac{1}{\alpha}X_{k_n}(g+1,j)\right)$. We have

$$\pi(X_{k_n}(g+1,j)) - \pi(x_{k_n}(g,j)) \le v_g^j(n_{k_n}) \le \pi(X_{k_n}(g+1,j)).$$

Apply the prime number theorem and divide by $n_k = n_{k_n}$ to obtain

$$\frac{v_g^j(n_k)}{n_k} = \alpha + o(1)\,,$$

as required. ∎

Problem 1. Is $\frac{v_g^j(n)}{n}$dense in a non trivial interval when $f(x) = \log(x)$ and N is the set of primes?

Next note that for the $f(x)$ of theorem 2.2 we have the following properties:

---

[1]Thanks are offered to an unknown referee for help with the proof of this theorem.

1. $f(x)$ increases to ∞ as $x \to \infty$.

2. $f'(x)$ decreases to 0 as $x \to \infty$.

3. $xf'(x) \to 0$ as $x \to \infty$.

If we strengthen (3) to

(3′) $xf'(x) = o\left(\frac{1}{\log(x)}\right)$,

a more general theorem holds for functions satisfying $(1), (2), (3')$. Two preliminary results are needed. The first was posed as a problem by Szegö [12] and was solved by Veress [16]: let $a_n > 0$, $\lim_{n\to\infty} a_n = 0$ and $\sum_{n=1}^{\infty} a_n$ divergent. If $s_n := \sum_{i=1}^{n} a_i$ then the $\{s_n\}$, $n = 1, 2, 3, \ldots$ are dense in [0,1]. The second is that $p_{n+1} - p_n = O(p_n)$ (Bertrand Postulate). Of course, the much stronger result due to Hoheisel says:

$$\pi(x + x^{1-\delta}) - \pi(x) \sim \frac{x^{1-\delta}}{\log(x)}\,,$$

for a constant $\delta > 0$. This implies that

$$p_{n+1} - p_n = O(p_n^{1-\delta}).$$

Although this is not needed for the proof of our theorem we make a few comments: The value of the constant is of great interest. Hoheisel (1930) [5] obtained $1-\delta = 1 - \frac{1}{33000} + \epsilon$. This has been refined by many authors. Ingham (1937) [6] obtained $1-\delta = \frac{5}{8} + \epsilon$. For about a year Mozzochi (1986) [9] held the record with the value $1-\delta = \frac{11}{20} - \frac{1}{384} = .54739\ldots$. There have been several improvements since. The best result obtained on the Riemann Hypothesis is due to Cramér (1920) [1], who showed $p_{n-1} - p_n = O(\sqrt{p_n}\ \log(p_n))$ via a study of his function:

$$V(x) = \sum_{\gamma>0} e^{\rho z}\,,$$

where $\rho = \beta + i\gamma$ are the non trivial zeros of the Riemann zeta function (other simpler proofs are available using Landau truncated form of the Riemann von Mangoldt formula (see Iveć [7]). It is interesting to note that in 1993 Gonek [4] conjectured that

$$\sum_{0<\gamma\leq T} x^{i\gamma} \ll Tx^{-1/2+\epsilon} + T^{1/2}x^{\epsilon}\,.$$

To obtain this he used the heuristic of replacing γ_n by its asymptotic value $\frac{2\pi n}{\log(n)}$. The assumption of his conjecture gives

$$\psi(x+h) - \psi(x) = h + O(h^{1/2}x^{\epsilon}),$$

for $1 \le h \le x$ and $\epsilon > 0$ ($\psi(x) = \sum_{p^k \le x} \log(p)$). This implies $p_{n+1} - p_n = O(p_n^{\epsilon})$. The latter was first conjectured by Piltz in 1884.

We now return to our theorem. The proof is a correction and variation from the case $f(n)$ to the case $f(p)$ of [11] pp. 284-285.

Theorem 2.3. *Let $f(x)$ have properties (1), (2), (3′) stated above. If $g(x)$ is Riemann integrable and $g(x)$ has a jump discontinuity at c and if $x_n = \{f(p_n)\}$ then the sequence*

$$\frac{g(x_1) + \ldots + g(x_n)}{n}, \; n = 1, 2, 3, \ldots$$

is dense in the interval determined by $g(c-)$ and $g(c+)$.

Proof. Let $N(x) = [f^{-1}(x)]$. Note that $f^{-1}(x)$ increases to ∞ and $(f^{-1})'(x) \to \infty$. Also,

$$\frac{f^{-1}(x-\epsilon)}{f^{-1}(x)} = o\left(\frac{1}{\log(f^{-1}(x))}\right), \; \forall\, \epsilon > 0$$

(use the mean value theorem:

$$\frac{f^{-1}(x-\epsilon)}{f^{-1}(x)} = \frac{f^{-1}(x-\epsilon)}{(f^{-1})'(c(x))\epsilon + f^{-1}(x-\epsilon)}$$

$$= \frac{f^{-1}(x-\epsilon) f'(f^{-1}(c(x))}{\epsilon + f^{-1}(x-\epsilon) f'(f^{-1}(c(x))}, \quad c(x) \in [x-\epsilon, x]$$

and the last expression is $o\left(\frac{1}{\log(f^{-1}(x))}\right)$ by (3′)). It follows that

$$\frac{N(x-\epsilon)}{N(x)} = o\left(\frac{1}{\log(f^{-1}(x))}\right), \; \forall\, \epsilon > 0.$$

By Veress Theorem the set x_n, $n = 1, 2, 3\ldots$ is dense in [0,1]. To see this write $f(p_n) = f(p_1) + (f(p_2) - f(p_1)) + \ldots + (f(p_n) - f(p_{n-1}))$. The terms of this sum go to zero by the mean value theorem, Bertrand Postulate and property (3′). Now let c be a point of continuity for $g(x)$. Abusing notation let x_n be a subsequence of x_n which converges to c. In this case the Cesàro means $\frac{g(x_1)+\ldots+g(x_n)}{n} \to g(c)$. If we have a jump at c and $x_n \to c$ let$|\, c - x_n \,| < \frac{\delta}{2}$. Note that $f(p_{n_0}) = [f(p_{n_0})] + \{f(p_{n_0})\} \leq [f(p_{n_0})] + c - \delta$ if $x_{n_0} < c - \delta$. It follows that $n_0 < p_{n_0} \leq N([f(p_{n_0})] + c - \delta)$. Observe that

$$\mu(\delta)\frac{n - N([f(p_{n_0})] + c - \delta)}{n} - M\frac{N([f(p_{n_0})] + c - \delta)}{n}$$
$$< \frac{g(x_1) + \ldots + g(x_{n_0}) + \ldots + g(x_n)}{n} < M\frac{N([f(p_{n_0})] + c - \delta)}{n} + M(\delta)\,,$$

where M is a bound on$\mid g(x) \mid$ and

$$\mu(\delta) = \inf_{|x-c|<\delta}(g(x)) \quad \text{and} \quad M(\delta) = \sup_{|x-c|<\delta}(g(x))$$

and x_{n_0} is the last of the x_n which is outside $\mid x - c \mid < \delta$ (in fact assume that $x_{n_0} < c - \delta$: The other possibility works even more easily). Note that $(1 + \epsilon)\log(n)n > p_n > N([f(p_n)] + c - \frac{\delta}{2})$ for large n by the prime number theorem and since $\mid c - x_n \mid < \frac{\delta}{2}$. It follows that:

$$\frac{N([f(p_{n_0})] + c - \delta)}{n} < \frac{N([f(p_n)] + c - \delta)(1 + \epsilon)\log(n)}{N([f(p_n)] + c - \frac{\delta}{2})}$$

$$= o\left(\frac{(1 + \epsilon)\log(n)}{\log(f^{-1}([f(p_n] + c - \frac{\delta}{2}))}\right) = o\left(\frac{(1 + \epsilon)\log(n)}{\log(n\log(n))}\right).$$

The last expression is $o(1)$ so, we conclude that the limit points of $\frac{g(x_1)+\ldots+g(x_n)}{n}$ are between $g(c-)$ and $g(c+)$ when $x_n \to c$. To see that the limit points for the full sequence cover the entire interval recall that $g(x)$ is continuous almost every where so, that there are points of continuity of $g(x)$ arbitrarly close to c. Choose points of continuity c' and c'' with $0 < c' < c < c'' < 1$ and subsequences $x_{n'}$ and $x_{n''}$ of x_n such that $x_{n'} \to c'$ $x_{n''} \to c''$. Let $g(x_1) + \ldots + g(x_n) = nG_n$. We have:

$$G_{n'} \to g(c') \text{ and } G_{n''} \to g(c'').$$

These values can be made to be with in ϵ of $g(c-)$ and $g(c+)$. For simplicity assume $g(c-) < g(c+)$ and observe

$$\mid G_n - G_{n+1} \mid = \left|\frac{G_n}{n+1} - \frac{g(x_{n+1})}{n+1}\right| < \frac{2M}{n+1}.$$

Let $\frac{g(c+)-g(c-)}{l} = \epsilon$ and choose n_0 so that $\frac{2M}{n_0+1} < \epsilon$. Partition $\mathbb{R}$ by

$$g(c-) - \epsilon < g(c-) < g(c-) + \epsilon < \ldots < g(c+) - \epsilon < g(c+) < g(c+) + \epsilon.$$

Choose $n_1 > n_0$ with G_{n_1} in one of the first two intervals and choose $n_2 > n_1$ with G_{n_2} in one of the last two intervals. In this case there is at least one k with $n_1 < k < n_2$ and G_k in any other (relevent) chosen interval of the partition of $\mathbb{R}$ (use contradiction). $\square$

To see that this Theorem implies that the $\frac{v_g^j(n)}{n}$ are dense in [0,1] for functions satisfying (1),(2) and (3′) apply it to $10^{j-1}f(x)$ and $g(x) = 1$ if $\frac{g}{10} \leq x < \frac{g+1}{10}$ and 0 otherwise.

Problem 2. Can the condition $xf'(x) = o\left(\frac{1}{\log(x)}\right)$ be relaxed to $xf'(x) = o\left(\frac{1}{(\log(x))^\beta}\right)$, $\beta \in (0,1)$ or perhaps $xf'(x) = o(1)$ in Theorem 2.3?

References

[1] H. Cramér, Some theorems concerning prime numbers, *Ark. Mat. Astron. Fysik.*, **15** (1920), 1–33.

[2] M. Drmota, R.F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics, **1651** (1997).

[3] J. Franel, Apropos de tables de logarithms, *Vjschr. Naturf. Ges. Zurich*, **62** (1917), 286–295.

[4] S. Gonek, An explicit formula of Landau and its applications to the theorey of the zeta-function, In: *A Tribute to Emil Grosswald: Number Theory and Related Analysis* (Ed-s Knopp, Seigorn), Contemporary Math., **143** (1993), 395–413.

[5] G. Hoheisel, Primzahlprobleme in der Analysis, *Sitzungsberichte Berliner Akad. D. Wiss.* (1930), 580–588.

[6] A.E. Ingham, On the difference of consecutive primes, *Quart. J. Pure and Appl. Math.*, Ser. 2, **8** (1937), 255–266.

[7] A. Iveć, *The Riemann Zeta-Function*, John Wiley and Sons (1985).

[8] L. Kuipers, H. Niederreiter, *Uniform Distribution of Sequences* , Wiley Interscience Publ. (1974).

[9] C.J. Mozzochi, On the difference between consecuative primes, *J. Nb. Th.* (1986), 181–187.

[10] H. Poincaré, *Science and Hypothesis*, Dover Publications (1952).

[11] G. Pólya, G. Szegö, *Problems and Theorems in Analysis 1*, Springer (1978)

[12] G. Szegö, Problem, *Arc. Math. Phys. Ser 3*, **23** (1914), 361.

[13] P. Turán, *A New Method of Analysis and its Applications*, Wiley Interscience Publ, (1985).

[14] H. Weyl, Über die Gleichverteilung von Zalhen mod Eins, *Math. Ann.*, **77** (1917), 313–352.

[15] A. Wintner, On the cyclical distribution of the logarithms of prime numbers, *Q. J. Math.*, **1** (1935), 65–68.

[16] P. Veress, *Arch. Math. Phy.*, Ser. 3, **25** (1917), 88.