

a^n の先頭桁の問題に対する χ^2 検定の応用

四丸直人・高嶋恵三

岡山理科大学理学研究科応用数学専攻

岡山理科大学理学部応用数学科

(2014年9月22日受付、2014年11月6日受理)

1 導入と背景

自然数 a に対して 10 進法で a^n を表示する場合、 a が 10 のべき乗でない限り a^n の先頭桁の数字には、1 から 9 までの数字が出現する。 a^n の先頭桁の数字を自然数 k ($1 \leq k \leq 9$) とすると、

$$k \times 10^{\ell-1} \leq a^n < (k+1) \times 10^{\ell-1}$$

が成立する。これは、

$$\log_{10} k \leq n \log_{10} a \pmod{1} < \log_{10}(k+1)$$

と同値である。ここで、 $x \pmod{1}$ は実数 x の小数部分を表す (cf. [1])。H.Weyl ([8]) により、無理数列 $\{n \log_{10} a \pmod{1}\}_{n>0}$ は漸近的に $[0, 1)$ 上一様分布することが知られている。これと上の不等式より、次の結果

$$\frac{\#\{m; a^m \text{の先頭桁} = k, m \leq n\}}{n} \xrightarrow{n \rightarrow \infty} \log_{10}(k+1) - \log_{10} k$$

が容易に得られる。ここで、 $\#A$ は集合 A の要素の数を表す。この事実より、先頭桁の数字の経験分布と極限分布の「近さ」を χ^2 検定で測る問題が考えられ、高嶋・小谷 ([6]) の報告がある。彼らは $a = 7$ の場合について一つの peak を持った「丘」が繰り返されることを報告している。

本報告では高嶋・小谷 ([6]) の議論を発展させ、 $\log_{10} a$ の連分数展開との関係を明らかにし、さらに、 $a = 7$ だけでなく $a = 33$, $a = 54$ の場合についても、 $a = 7$ の場合とは異なる形の「丘」が繰り返されることを報告し、それらの数値計算の結果について報告する。その際、この「丘」がその高さを減じながら繰り返し出現するという現象は、 $\log_{10} a$ の連分数展開の中に出現する、大きな孤立した部分分母 a_{η} に深く関係し、その部分分母までによる近似分数 p_{η}/q_{η} の分母の大きさ q_{η} が、繰り返しの「周期」にほぼ等しいこと、および、その大きな孤立した部分分母の 1 つ前までの近似分数 $p_{\eta-1}/q_{\eta-1}$ の分母 $q_{\eta-1}$ の step 毎に χ^2 検定の値の変化を見ることにより、グラフの形の概略を見ることができると報告する。

2 無理数回転と $\log_{10} a$ の連分数展開について

上記の述べたように、 a^n の先頭桁の数字の問題は、 $n \log_{10} a$ の挙動の問題でもある。実数 $\alpha (> 0)$ による無理数回転 $R_{\alpha}(x) = x + \alpha \pmod{1}$ を考える。また、実数 $\alpha (< 1)$ の連分数展開を Drmota and Tichy ([2])

に従って、以下のように表記することにする。

$$\alpha = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \frac{1}{a_5 + \frac{1}{a_6 + \frac{1}{a_7 + \frac{1}{a_8 + \frac{1}{a_9 + \dots}}}}}}}}}}$$

このとき、実数 α は以下のようにして定まる有理数 $r_n = \frac{p_n}{q_n}$ によって非常によく近似されることが分かっている (cf. [3]) :

$$p_0 = 0, p_1 = 1, p_{n+1} = a_{n+1}p_n + p_{n-1},$$

$$q_0 = 1, q_1 = a_1, q_{n+1} = a_{n+1}q_n + q_{n-1},$$

$$|\alpha - r_n| < \frac{1}{q_n q_{n+1}}.$$

特に、 $a = 7$ の場合、 $\log_{10} 7$ の連分数展開は以下の通りである :

$$\log_{10} 7 = \frac{1}{1 + \frac{1}{5 + \frac{1}{2 + \frac{1}{5 + \frac{1}{6 + \frac{1}{1 + \frac{1}{4813 + \frac{1}{1 + \frac{1}{1 + \dots}}}}}}}}}}$$

この場合のように、連分数展開の部分分母 a_n に、非常に大きな値が出現する場合は (上記の例では $a_7 = 4813$)、 $q_6 = 510$, $q_7 = 2455069$, $q_8 = 2455579$ となり、無理数回転 $R_\alpha^n = n\alpha \pmod{1}$ ($\alpha = \log_{10} 7$) は有理数回転 $n \times r_6 = n \times \frac{431}{510} \pmod{1}$ で非常によく近似されることが分かる :

$$|\log_{10} 7 - r_6| = \left| \log_{10} 7 - \frac{431}{510} \right| < \frac{1}{510 \times 2455069}.$$

以下では、このような孤立した、大きな部分分母を a_η と表すことにする。即ち、 $\eta = 7$ である。

3 χ^2 検定とその数値計算の結果

a^n の先頭桁の数字の経験分布とその極限分布との差を測る道具として、統計学で「適合度の検定」として有名な χ^2 検定を考える。

数字 k ($k = 1, 2, \dots, 9$) の観測度数 $O_k(n) = O_k$ は次のように定義される :

$$O_k = \#\{m; a^m \text{ の先頭桁の数字} = k, 1 \leq m \leq n\}.$$

また、期待度数 $E_k(n) = E_k$ は

$$E_k = n \times (\log_{10}(k+1) - \log_{10} k) = n \times \log_{10} \left(1 + \frac{1}{k} \right)$$

で与えられる。この時、 χ^2 検定の検定値 $\chi_0^2(n) = \chi_0^2$ は次で与えられる：

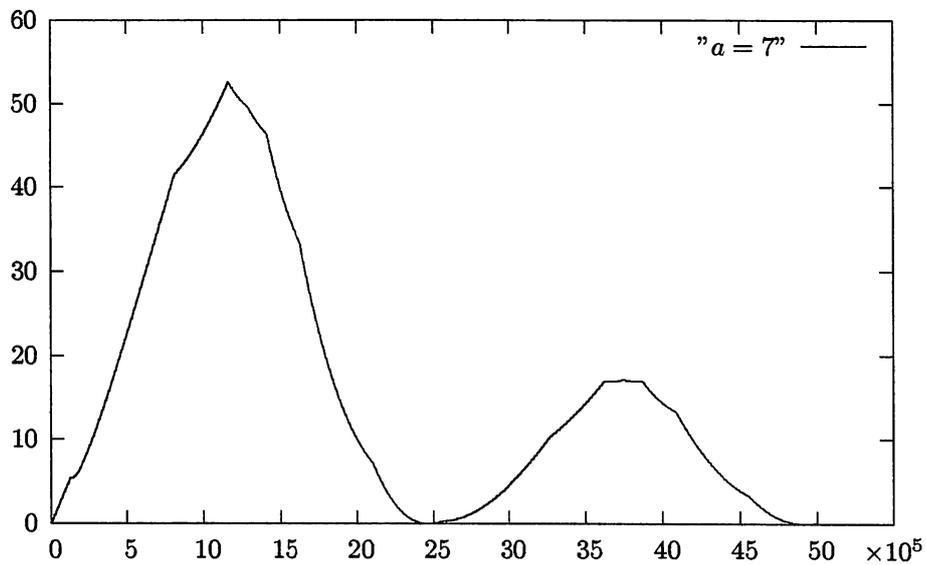
$$\chi_0^2(n) = \chi_0^2 = \sum_{k=1}^9 \frac{(O_k(n) - E_k(n))^2}{E_k(n)} = \sum_{k=1}^9 \frac{(O_k - E_k)^2}{E_k} .$$

この χ^2 検定の自由度は $8(=9-1)$ である。

3.1 $a = 7$ の場合の数値計算結果

Fig.1 に示されるように、 $a = 7$ の場合、グラフ上、一つの peak を持つ「丘」の繰り返しが見られ、それぞれの「谷」の間隔は $q_7 = 2455069$ であり、最初の谷 $q_7 = 2455069$ と $q_8 = 2455579$ の間は、「やや広い」谷間になっており、その幅は $q_6 = 510$ に等しくなっている。高嶋・小谷 ([6]) で報告されたように、このような一定の幅 ($q_7 = 2455069$) を持った「丘」と「谷」が暫くは繰り返される。ここで、 $\log_{10} 7$ の $r_6 = \frac{431}{510}$ の近似の議論から、 χ^2 検定値のグラフを $\text{step } q_6 = 510$ で表示することの重要性が分かる。また、数値計算

Fig.1 $a = 7, n = 5,000,000, \text{step} = 510$

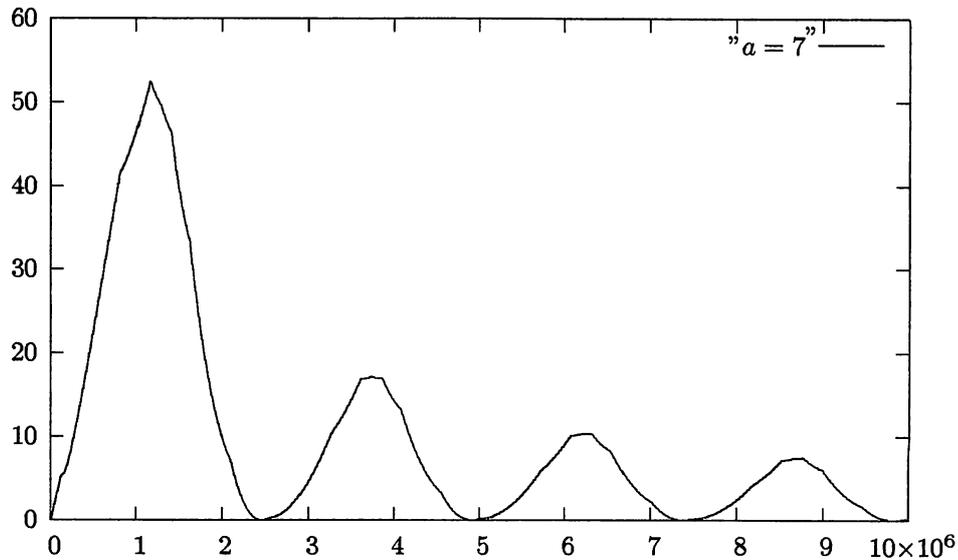


の結果では $\chi_0^2(n)$ の最大値は n が約 1,200,00 程の所で、その値は 52 以上となっている。自由度 8 の χ^2 検定でその検定値が 52 を超える確率は概ね 100,000,000 分の 2 程度である。

3.2 $a = 33$ と $a = 54$ の場合

まず、 $a = 33$ の場合、 $\log_{10} 33 - 1$ の連分数展開は以下のようになる：

$$\log_{10} 33 - 1 = \frac{1}{1 + \frac{1}{1 + \frac{1}{13 + \frac{1}{299 + \frac{1}{1 + \frac{1}{1 + \frac{1}{10 + \frac{1}{1 + \frac{1}{14 + \dots}}}}}}}}}} .$$

Fig.1a $a = 7$, $n = 10,000,000$, $\text{step} = 510$ 

この展開の結果より、 $a_4 = 299$, $r_3 = \frac{14}{27}$, $r_4 = \frac{4187}{8075}$, $q_3 = 27$, $q_4 = 8075$ であり、 $a = 7$ の場合と同様の議論が有効であることが分かる。しかし、Fig.2 で明示されているように $a = 33$ の場合のグラフは、明らかに 2つの peak を持った「丘」の繰り返しである。このことは $a = 7$ の場合と明確な差異であるが、今後の課題とする。

次に、 $a = 54$ の場合は $\log_{10} 54 - 1$ の連分数展開は以下ようになる：

$$\log_{10} 54 - 1 = \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{3 + \frac{1}{1 + \frac{1}{326 + \frac{1}{2 + \dots}}}}}}}}}} .$$

これより、 $q_7 = 71$, $a_8 = 326$, $q_8 = 23202$ であり、上記の $a = 7$, $a = 33$ の場合と同様の議論が適用できることが分かる。然しながら、Fig.3 に明らかなように、 χ_0^2 の値のグラフの概略はいずれの場合とも明確に異なることも分かる。

これらの例から分かるように、 a^n の先頭桁の数字の経験分布の漸近挙動を考える場合、 χ_0^2 の値を $\text{step } q_{\eta-1}$ 毎に考えることでグラフの概略を取り出すことができ、問題の本質的な差異を明確にできることが分かる。但し、ここで a_η は $\log_{10} a$ の連分数展開における非常に大きな部分分母であり、例えば、 $\log_{10} 7$ の場合、 $\eta = 7$, $\log_{10} 33$ の場合、 $\eta = 4$, $\log_{10} 54$ の場合、 $\eta = 8$ である。

連分数展開の一般理論より

$$r_2 < r_4 < \dots < r_{2n} = \frac{p_{2n}}{q_{2n}} < \dots < \alpha < \dots < r_{2n-1} = \frac{p_{2n-1}}{q_{2n-1}} < \dots < r_5 < r_3,$$

である。このことより無理数回転 $n \times \alpha$ の有理数回転については η が奇数であるか ($a = 7$ の場合)、偶数であるか ($a = 33$, $a = 54$ の場合) により近似の仕方が異なるが、議論そのものは自明な修正により同様に行えることも分かる。また、Fig.1, Fig.2, および Fig.3 より明らかに次のことが分かる：

Fig.2 $a = 33$, $n = 16,000$, $\text{step} = 27$

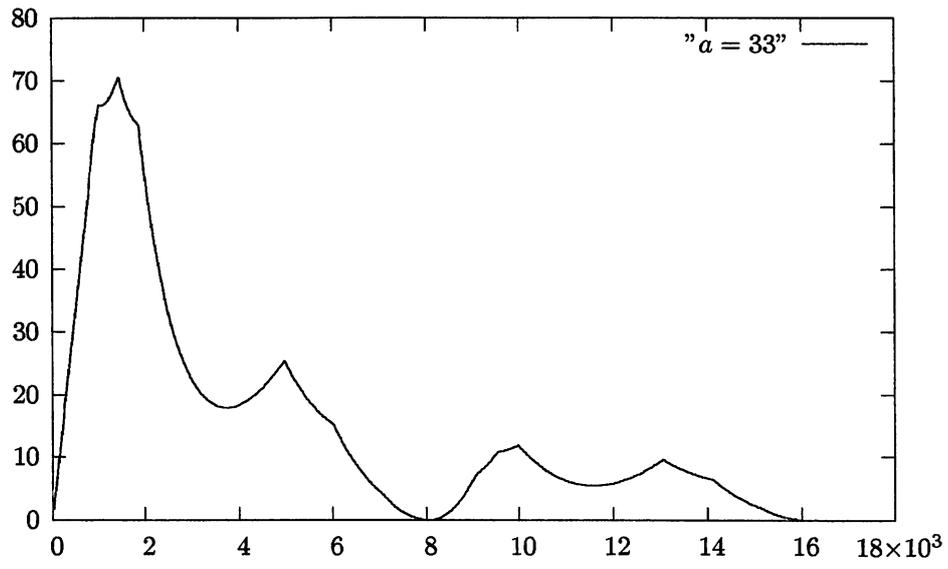
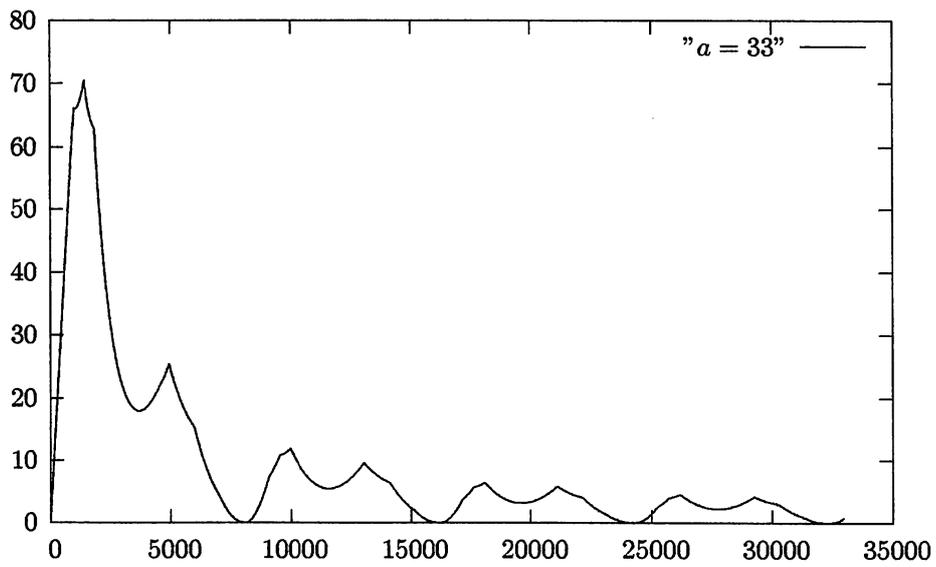


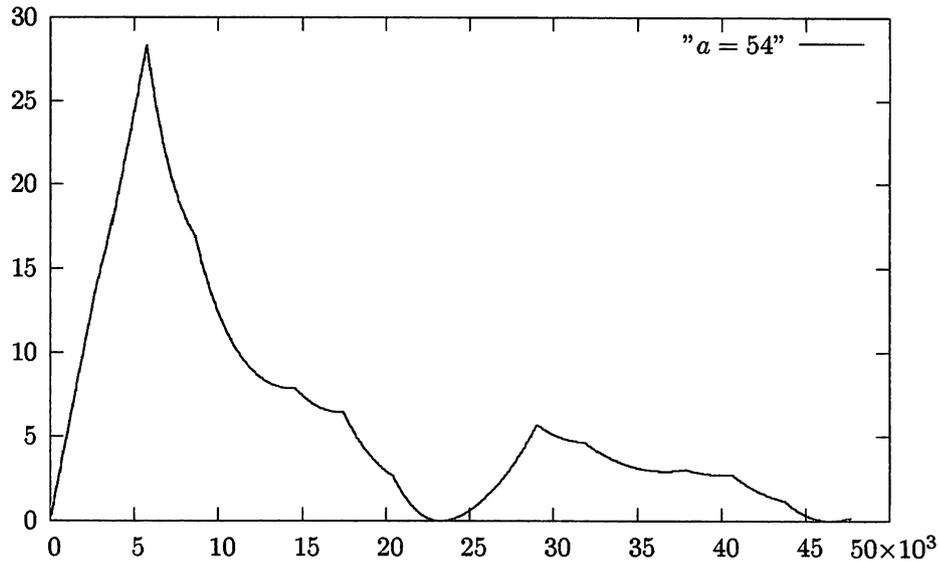
Fig.2a $a = 33$, $n = 33,000$, $\text{step} = 27$



$a = 7$ の場合は「丘」は1つの peak を持つが、 $a = 33$ の場合には「丘」は2つの peak を持つ。さらに、 $a = 54$ の場合には $a = 7, a = 33$ のいずれの場合とも異なる形の「丘」が繰り返される。

これらの議論の詳細については Mori and Takashima ([5]) に譲るが、そこでは、 $\log_{10} a$ の連分数展開の中に、孤立した大きな部分分母の出現することが重要であるだけでなく、各 $\log_{10} k$ ($k = 1, 2, \dots, 9$) と近似分数 $r_{\eta-1}$ の分母 $q_{\eta-1}$ から派生する分数 $j/q_{\eta-1}$ との関係が重要であることが示されている。これらの考察

Fig.3 $a = 54, n = 50,000, \text{step} = 71$

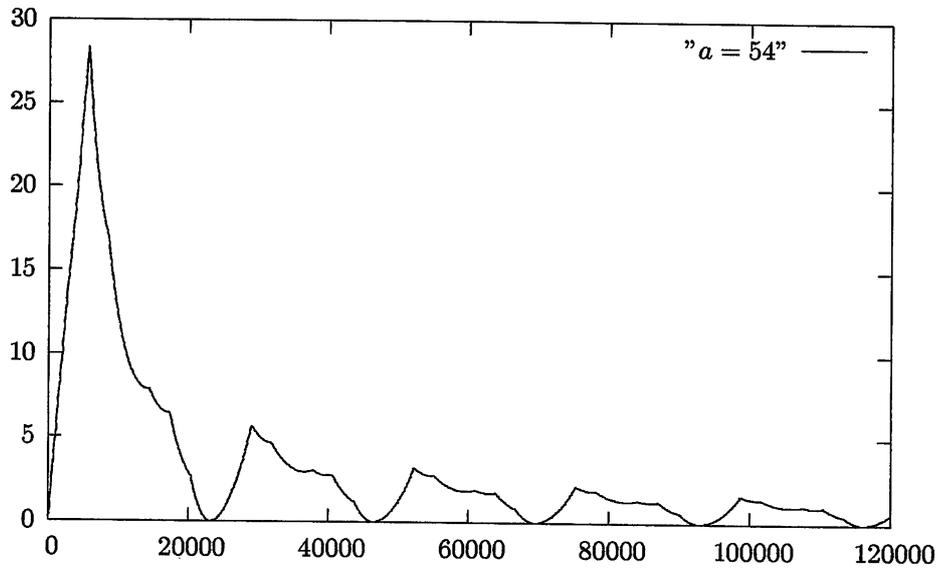


より、どの様な超越数 $\log_{10} a$ に対して、どのような形の「丘」が出現するのか、という問題が生ずるが、超越数 $\log_{10} k$ ($k = 1, 2, \dots, 9$ と $\log_{10} a$ の近似分数との関係が基本的な問題であるので極めて難しい問題であるといえ、今後の問題である。

References

- [1] Berger, A.: *Chaos and Chance*, Walter de Gruyter, Berlin, 2001.
- [2] Drmota, M. – Tichy, R. F.: *Sequences, Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer-Verlag, Berlin, 1997.
- [3] Hardy, G. H. – Wright, E. M.: *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, Oxford, 1979.
- [4] Khinchin, A. Ya.: *Continued Fractions*, Dover Publications, New York, 1997.
- [5] Mori, Y. – Takashima, K.: *On χ^2 tests of leading digits of a^n , with $\log_{10} a$ having single isolated large partial quotient*, to be submitted to *Uniform Distribution Theory*.
- [6] 高嶋恵三、小谷真美: べき乗の先頭桁の数字について、岡山理科大学紀要, **42 A** (2006), 7–11.
- [7] 高嶋恵三、長濱紗智、林紘: べき乗の先頭桁、無理数回転および連分数展開, 岡山理科大学紀要, **44 A** (2008), 9–13.

Fig.3a $a = 54$, $n = 120,000$, $\text{step} = 71$



[8] Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.*, **77** (1916), 313–352.

Application of χ^2 test to the problem of leading digits of a^n

Shimaru, N, and Takashima, K.

*Department of Applied Mathematics,
Faculty of Science
Okayama University of Science
1-1, Ridai-cho, kita-ku, Okayama 700-0005, Japan*

(Received September 22, 2014; accepted November 6, 2014)

In this report, we discuss with the problem of leading digits of a^n , where a is a natural number, which is not equal to any power of 10. We report that in case that the common logarithm of a , $\log_{10} a$, has single isolated large partial quotient, graphs of χ^2 tests of the goodness of fits between empirical distributions of leading digit k 's and their limit distributions show very strange shapes and repetitions of "hills" and "valleys".

We give some explanations for these unusual phenomena by considering the continued fraction expansions of $\log_{10} a$ and their approximations by "rational numbers", derived their continued fraction expansions. We also report there are some different types of shapes of graphs, at least 3 types (cf. Fig.1 , fig.2 and fig.3), with one "peak" (in case of $a = 7$), with "two peaks" (in case of $a = 33$), and another shape (in case of $a = 53$).

Keywords: *leading digit, irrational rotation, χ^2 test, continued fraction expansion*