SOME RESULTS ON THE DISTRIBUTION OF THE MOST SIGNIFICANT DIGIT

By A. K. ADHIKARI

Indian Statistical Institute

SUMMARY. This paper finds the distribution of the most significant digit of some functions of random variables $X_1, X_2, \ldots, X_n$, where these variables are independent and distributed uniformly in (0, 1). The probability that the most significant digit of Y_n is A $(A = 1, \ldots, 9)$ has been found, where Y_n is defined as the product of reciprocals of n such random variables. It has been shown that this probability distribution tends to $\log_{10} (A+1)/A$ as n tends to infinity. Similarly if Z_n is defined as $Z_n = X_1/X_2/\ldots/X_{n+1}$, it has been proved that the probability distribution of the most significant digit of Z_n also tends to the same limit as n tends to infinity. More generally it is found that if $V_1, V_2, \ldots, V_n$ are defined as $V_1 = B/X_1, \ldots,\ V_n = V_{n-1}/X_n$ where B is any random variable defined on the positive axis of the real line, the probability distribution of the most significant digit tends to $\log_{10} (A+1)/A$ as n tends to infinity.

Benford (1938) observed in statistical tables that more entries start with smaller significant digits. The proportion of entries with most significant digit $A(A = 1, 2, \ldots, 9)$ is approximately $\log_{10}(A+1)/A$. His findings created interests among many persons and they tried to find the reasons for such an 'abnormal law' (see, Furry and Hurtwitz, 1945 ; Goudsmit and Furry, 1944 ; Pinkham, 1961; Flehinger, 1966).

In an earlier paper (Adhikari and Sarkar, 1968) the distribution of the most significant digits were found for some functions of random variables distributed uniformly in (0, 1). It was shown that if $g_{1,n}$ and $g_{2,n}$ are defined as $g_{1,n} = X_1^n$ and $g_{2,n} = X_1 . X_2 \ldots X_n$ where $X_1, X_2, \ldots, X_n$ are all independent and uniformly distributed random variables from (0, 1), the probability that the most significant digit of $g_{1,n}$ is A tends to $\log_{10} (A+1)/A$, as n tends to infinity, and the same is true for $g_{2,n}$. In the present paper the author finds the distribution of the most significant digit of some other functions of these random variables.

For our convenience, unless otherwise mentioned, we shall use the same symbol f to denote the probability density function of any random variable. We shall also use the following symbols :

A : a positive integer lying between 1 to 9;

m.s.d. (y) : most significant digit of y;

$\ln x = \log_e x$;

$$q(t) = \frac{(A+1)^{1-t} - A^{1-t}}{A^{1-t} . (A+1)^{1-t}} . \frac{10^{1-t}}{10^{1-t}-1} ;$$

X_i = uniformly distributed random variable in (0, 1), which is independent to any other $X_j\ (j \neq i)$ i.e. the probability density function of X_i is given by

$$f(x_i) = \begin{cases} 1 \text{ if } 0 < x_i < 1 \\ 0 \text{ otherwise} \end{cases}$$

Y_n = a random variable defined as $\frac{1}{X_1} \cdot \frac{1}{X_2} \cdots \frac{1}{X_n}$

Z_n = a random variable defined by the recursive relation $Z_n = Z_{n-1}/X_{n+1}$

$U(0, 1)$ = uniform distribution in (0, 1).

We shall first find the distribution of the random variable Y_n. It can be easily seen that the density function f_n of Y_n is given by

$$f_n(y) = \begin{cases} \frac{(\ln y)^{n-1}}{(n-1)!\,y^2} & \text{if } 1 \leqslant y < \infty \\ 0 & \text{otherwise.} \end{cases} \quad \ldots \quad (1)$$

The result can be seen to be true for $n = 1$ and 2. Using the method of induction (1) can be established since the joint density function of Y_r and $1/X_{r+1} = W$ is given by

$$f(y, w) = \begin{cases} \frac{1}{(r-1)!} \frac{(\ln y)^{r-1}}{y^2} \frac{1}{w^2} & \text{if } y \geqslant 1 \text{ and } w \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

if (1) holds for $n = 1, 2, \ldots, r$.

Hence the joint density function of the transformed variables $Y_{r+1} = Y_r . W$ and $T = W$, is given by

$$f((y, t) = \begin{cases} \frac{1}{(r-1)!} \frac{1}{y^2} \frac{(\ln y - \ln t)^{r-1}}{t} & \text{for } 1 \leqslant t \leqslant y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

From above, the marginal distribution of Y_{r+1} can be found by integrating $f(y, t)$ with respect to t in the range 1 to y, which yields the result given in (1).

Define

$$p_n(A) = \text{Prob}\,\{\text{m.s.d.}\,(Y_n) = A\}.$$

The m.s.d. (Y_n) will be A if Y_n lies in the interval $[A, A+1)$ or $[10A, 10(A+1))$ etc. i.e. if Y_n lies in any of the interval $[A.10^r, (A+1)\,10^r)$ for $r = 0, 1, \ldots, \infty$. Hence

$$p_n(A) = \sum_{r=0}^{\infty} \frac{1}{(n-1)!} \int_{A \cdot 10^r}^{(A+1)10^r} \frac{(\ln y)^{n-1}}{y^2}\, dy$$

$$= \frac{1}{(n-1)!} \sum_{r=0}^{\infty} \Big[-\frac{1}{y} \{(\ln y)^{n-1} + (n-1)(\ln y)^{n-2} + \ldots + (n-1)!\} \Big]_{A.10^r}^{(A+1)10^r}$$

$$= \sum_{r=0}^{\infty} \left[-\frac{1}{y} \left\{ 1+\ln y+\ldots+\frac{(\ln y)^{n-1}}{(n-1)!} \right\} \right]_{A.10^r}^{(A+1)10^r}$$

$$= \sum_{r=0}^{\infty} \left[-\frac{1}{y} \times \text{coefficient of } t^{n-1} \text{ in } \frac{e^{t \ln y}}{1-t} \right]_{A.10^r}^{(A+1)10^r} \quad \text{for } |t| < 1$$

$$= -\sum_{r=0}^{\infty} \left[\text{ coefficient of } t^{n-1} \text{ in} \frac{e^{-\ln y(1-t)}}{1-t} \right]_{A.10^r}^{(A+1)10^r}$$

$$= \text{coefficient of } t^{n-1} \text{ in } \frac{q(t)}{1-t} \quad \text{for } |t| < 1, \text{ where}$$

$$q(t) = \frac{(A+1)^{1-t}-A^{1-t}}{[A.(A+1)]^{1-t}} \cdot \frac{10^{1-t}}{10^{1-t}-1}. \qquad \ldots \quad (2)$$

Using the same method as has been done in the earlier paper (Adhikari and Sarkar, 1968) we can find $\lim_{n \to \infty} p_n(A)$ here.

Let

$$q(t) = a_0+a_1t+a_2t^2+\ldots$$

then

$$p_n(A) = a_0+a_1+\ldots+a_{n-1}.$$

Hence

$$\lim_{n \to \infty} p_n(A) = a_0+a_1\ldots+a_n+\ldots$$

$$= \lim_{t \to 1} q(t).$$

But

$$\lim_{t \to \infty} q(t) = \lim_{t \to 1} \frac{10^{1-t}}{10^{1-t}-1} \cdot \frac{(A+1)^{1-t}-A^{1-t}}{A^{1-t}.(A+1)^{1-t}}$$

$$= \frac{\ln \dfrac{A+1}{A}}{\ln 10} = \log_{10} \frac{A+1}{A}.$$

Hence we have the theorem :

Theorem 1 : *If* Y_n *is defined as* $Y_n = \frac{1}{X_1} \cdot \frac{1}{X_2} \cdots \frac{1}{X_n}$ *where* $X_1, X_2, \ldots, X_n$ *are all independent and identically distributed random variable in* $U(0, 1)$, *then*

(i) *the probability density function of* Y_n

$$f_n(y) = \begin{cases} \dfrac{(\ln y)^{n-1}}{(n-1)y^2} & \text{if } 1 \leqslant y < \infty \\ 0 & \text{otherwise} \end{cases}$$

and for $A = 1, 2, \ldots, 9$;

(ii) *probability* {*the m.s.d. of* $Y_n = A$} $= p_n(A)$

$$= \sum_{r=0}^{\infty} \left[\frac{-1}{y} \left\{ 1 + \ln y + \ldots + \frac{(\ln y)^{n-1}}{(n-1)} \right\} \right]_{A.10^r}^{(A+1)\cdot 10^r}$$

$$= \text{coefficient of } t^{n-1} \text{ in } \frac{(A+1)^{1-t} - A^{1-t}}{A^{1-t}(A-1)^{1-t}} \cdot \frac{10^{1-t}}{10^{1-t}-1} \cdot \frac{1}{1-t} \quad \text{for } |t| < 1$$

and

(iii) $$\lim_{n \to \infty} p_n(A) = \log_{10} \frac{A+1}{A}.$$

Lemma 1 : *For any fixed integer n,* $\sum_{r=0}^{\infty} \frac{r^n}{10^r}$ *is bounded.*

Proof : For any fixed integer n, and $0 < x < 1$

$$\sum_{r=0}^{\infty} r^n x^r \leqslant \sum_{r=0}^{\infty} r(r+1)(r+2) \ldots (r+n-1)x^r.$$

But $$\sum_{r=0}^{\infty} r(r+1) \ldots (r+n-1)x^r = \frac{1}{(1-x)^{n+1}}.$$

Hence, $$\sum_{r=0}^{\infty} \frac{r^n}{10^r} \leqslant \frac{1}{\left(1-\frac{1}{10}\right)^{n+1}} = \frac{10^{n+1}}{10^{n+1}-1} < 1.5$$

for any positive integer n.

Lemma 2 : *For any positive integer n and for* $D = 1, 2, \ldots, 10$

$$\sum_{r=1}^{\infty} \frac{1}{D.10^r} \left[2 \ln (D.10^r) + \frac{\{2 \ln (D.10^r)\}^2}{2!} + \ldots + \frac{\{2 \ln (D.10^r)\}^{n-1}}{(n-1)!} \right] \leqslant B.(n-1),$$

where B is a finite number.

Proof : Let $\alpha = 2 \ln 10$ and

$$C_{n,r}(D) = \frac{1}{D.10^r} \left[2 \ln (D.10^r) + \frac{\{2 \ln (D.10^r)\}^2}{2!} + \ldots + \frac{\{2 \ln (D.10^r)\}^{n-1}}{(n-1)!} \right].$$

Since $D = 1, \ldots, 10$,

$$C_{n,r}(D) \leqslant \frac{1}{D.10^r} \left[2(r+1) \ln 10 + 2^2(r+1)^2 \frac{(\ln 10)^2}{2!} + \ldots + 2^{n-1}(r+1)^{n-1} \frac{(\ln 10)^{n-1}}{(n-1)!} \right]$$

$$= \frac{1}{D.10^r} \left[(r+1)\alpha + (r+1)^2 \frac{\alpha^2}{2!} + \ldots + (r+1)^{n-1} \frac{\alpha^{n-1}}{(n-1)!} \right]$$

$$\therefore \qquad \sum_{r=1}^{\infty} C_{n,r}(D) \leqslant \frac{1}{D}\left[\alpha \sum_{r=1}^{\infty} \frac{r+1}{10^r} + \frac{\alpha^2}{2!} \sum_{r=1}^{\infty} \frac{(r+1)^2}{10^r} + \ldots + \frac{\alpha^{n-1}}{(n-1)!} \sum_{r=1}^{\infty} \frac{(r+1)^{n-1}}{10^r}\right]$$

$$< \frac{15}{D}\left[\alpha + \frac{\alpha^2}{2!} + \ldots + \frac{\alpha^{n-1}}{(n-1)!}\right]$$

$$< \frac{15}{D} \cdot \frac{\alpha^4}{4!}(n-1),$$

since by Lemma 1 $\displaystyle\sum_{r=1}^{\infty} \frac{(r+1)^p}{10^r} = 10 \sum_{r=0}^{\infty} \frac{r^p}{10^r} < 15$

and $\alpha = 4.605.$

With the help of the above two lemmas we can prove the following theorem :

Theorem 2 : *Let $Z_1, Z_2, \ldots, Z_n$ be n random variables defined as $Z_1 = X_1/X_2$, $Z_2 = Z_1/X_3, \ldots, Z_n = Z_{n-1}/X_{n+1}$. Then*

(i) *the probability density function $f_n(z)$ of Z_n is given by*

$$f_n(z) = \begin{cases} \dfrac{1}{2^n} & \text{if} \quad 0 < z \leqslant 1 \\ \dfrac{1}{2^n z^2} + \dfrac{1}{2^{n-1} z^2} \displaystyle\sum_{j=1}^{n-1} 2^{j-1} \frac{(\ln z)^j}{j!} & \text{if} \quad 1 < z < \infty \\ 0 & \text{otherwise} \end{cases} \qquad \ldots (3)$$

(ii) *Prob {the m.s.d. of $Z_n = A$} $= q_n(A)$*

$$= \frac{1}{9.2^n} - \sum_{r=0}^{\infty}\left[\frac{1}{y}\left\{1 + \ln z + \ldots + \frac{(\ln z)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1).10^r}$$

$$+ \frac{1}{2^n} \sum_{r=0}^{\infty}\left[\frac{1}{y}\left\{1 + 2 \ln z + \ldots + \frac{(2 \ln z)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1)10^r} \qquad \ldots (4)$$

and

(iii) $$\lim_{n \to \infty} q_n(A) = \log_{10} \frac{A+1}{A}.$$

Proof : (i) It is easy to see that

$$f_1(z) = \begin{cases} \dfrac{1}{2} & \text{if } 0 < z \leqslant 1 \\ \dfrac{1}{2z^2} & \text{if } 1 < z < \infty \\ 0 & \text{otherwise} \end{cases}$$

and $f_2(z)$ is given by

$$f_2(z) = \begin{cases} \frac{1}{4} & \text{if } 0 < z \leqslant 1 \\ \frac{1}{4z^2} + \frac{\ln z}{2z^2} & \text{if } 1 < z < \infty \\ 0 & \text{otherwise} \end{cases}$$

which shows that (3) holds for $n = 1$ and 2. We use the method of induction to prove this. If (3) holds for $n = 1, 2, \ldots, r$ then the joint density function $f_r(z, x)$ of Z_r and X_{r+1} is given by

$$f_r(z. x) = \begin{cases} \frac{1}{2^r} & \text{if } 0 < z \leqslant 1 \text{ and } 0 < x < 1 \\ \frac{1}{2^r z^2} + \frac{1}{2^{r-1} z^2} \sum_{j=1}^{r-1} 2^{j-1} \frac{(\ln z)^j}{j!} & \text{if } z > 1 \text{ and } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence the joint density function $f_{r+1}(z, y)$ of the transformed variables $Z_{r+1} = Z_r/X_{r+1}$ and $Y = X_{r+1}$ is

$$f_{r+1}(z, y) = \begin{cases} \frac{y}{2^r} & \text{if } 0 < z \leqslant 1 \text{ and } 0 < y < 1 \text{ or} \\ & \text{if } z > 1 \text{ and } 0 < y \leqslant 1/z \\ \frac{1}{2^r z^2 y} + \frac{1}{2^{r-1} z^2 y} \sum_{j=1}^{r-1} 2^{j-1} \frac{(\ln zy)^j}{j!} & \text{if } z > 1 \text{ and } \frac{1}{z} < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence the marginal distribution of Z_{r+1} is

$$f_{r+1}(z) = \begin{cases} \frac{1}{2^{r+1}} & \text{if } 0 < z \leqslant 1 \\ \frac{1}{2^{r+1} z^2} + \frac{1}{2^r z^2} \sum_{j=1}^{r} 2^{j-1} \frac{(\ln z)^j}{j!} & \text{if } z > 1 \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Probability {the m.s.d. of $Z_n = A$}

$$= q_n(A)$$

$$= \frac{1}{2^n} \sum_{r=1}^{\infty} \int_{A/10^r}^{(A+1)/10^r} dy + \frac{1}{2^n} \sum_{r=0}^{\infty} \int_{A/10^r}^{(A+1)/10^r} \frac{1}{y^2} \left\{1+2\ln y+\ldots+\frac{(2\ln y)^{n-1}}{(n-1)!}\right\} dy$$

$$= \frac{1}{9.2^n} - \frac{1}{2^n} \sum_{r=0}^{\infty} \left[\frac{1}{y}\left\{(2^n-1)+(2^{n-1}-1)(2\ln y)+\ldots+\frac{(2\ln y)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1)10^r}$$

$$= \frac{1}{9.2^n} - \sum_{r=0}^{\infty} \left[\frac{1}{y}\{1+\ln y+\ldots+(\ln y)^{n-1}\}\right]_{A.10^r}^{(A+1).10^r}$$

$$+\frac{1}{2^n} \sum_{r=0}^{\infty} \left[\frac{1}{y}\left\{1+2\ln y+\ldots+\frac{(2\ln y)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1).10^r}$$

$$= P_n + Q_n(A) + R_n(A)$$

where $$P_n = \frac{1}{9.2^n}$$

$$Q_n(A) = -\sum_{r=0}^{\infty}\left[\frac{1}{y}\left\{1+\ln y+\ldots+\frac{(\ln y)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1).10^r}$$

and $$R_n(A) = \frac{1}{2^n}\sum_{r=0}^{\infty}\left[\frac{1}{y}\left\{1+2\ln y+\ldots+\frac{(2\ln y)^{n-1}}{(n-1)!}\right\}\right]_{A.10^r}^{(A+1).10^r}$$

(iii) $$\lim_{n\to\infty} q_n(A) = \lim_{n\to\infty} P_n + \lim_{n\to\infty} Q_n(A) + \lim_{n\to\infty} R_n(A)$$

But $$\lim_{n\to\infty} P_n = \lim_{n\to\infty}\frac{1}{9.2^n} = 0$$

and $Q_n(A)$ defined above can be seen to be same as $p_n(A)$ defined in Theorem 1. Hence $\lim_{n\to\infty} Q_n(A) = \log_{10}\frac{A+1}{A}$.

Again, $R_n(A)$ can be written as

$$R_n(A) = \frac{1}{2^n}\left[\frac{1}{A+1}\left\{1+2\ln(A+1)+\ldots+\frac{2^{n-1}(\ln A+1)^{n-1}}{(n-1)!}\right\}\right.$$

$$\left.-\frac{1}{A}\left\{1+2\ln A+\ldots+\frac{2^{n-1}(\ln A)^{n-1}}{(n-1)!}\right\}\right]+\frac{1}{2^n.9}\left(\frac{1}{A+1}-\frac{1}{A}\right)$$

$$+\frac{1}{2^n}\left[\sum_{r=1}^{\infty} C_{n,r}(A+1)-\sum_{r=1}^{\infty} C_{n,r}(A)\right]$$

where $$C_{n,r}(D) = \frac{1}{D.10^r}\left[2\ln(D.10^r)+\frac{\{2\ln(D.10^r)\}^2}{2!}+\ldots+\frac{\{2\ln(D.10^r)\}^{n-1}}{(n-1)!}\right].$$

In Lemma 2 it has been shown that $\sum_{r=1}^{\infty} C_{n,r}(D) < B.(n-1)$ where $B = \frac{15.(4.605)^4}{4!} < 300$, for all positive integer n and for $D = 1, 2, \ldots, 10$. Also it is clear that $\sum_{r=1}^{\infty} C_{n,r}(D) \geqslant 0$ for $n = 1, 2, \ldots$. Hence $\frac{1}{2^n}\sum_{r=1}^{\infty} C_{n,r}(A+1)$ and $\frac{1}{2^n}\sum_{r=1}^{\infty} C_{n,r}(A)$ both tend to zero as n tends to infinity. Hence $\lim_{n\to\infty} R_n(A) = 0$ and $\lim_{n\to\infty} q_n(A) = \log_{10}\frac{A+1}{A}$.

Corollary 1 : $\sum_{A=1}^{9} R_n(A) = -\frac{1}{2^n}$ *for all* $n = 1, 2, \ldots$.

Proof: For all n, $\sum_{A=1}^{9} q_n(A) = 1 = \sum_{A=1} \frac{1}{2^n.9}+\sum_{A=1}^{9} Q_n(A)+\sum_{A=1}^{9} R_n(A)$

$$= \frac{1}{2^n}+1+\sum_{A=1}^{9} R_n(A)$$

$$\therefore \sum_{A=1}^{9} R_n(A) = -\frac{1}{2^n}$$

It can be seen easily—by using the same method as earlier—that if u_n is defined as

$$u_n = b/X_1 X_2 \ldots X_n, \text{ then } \lim_{n \to \infty} \text{Prob \{the m.s.d. of } u_n = A\} = \log_{10} \frac{A+1}{A}.^*$$

From this the following theorem can be proved :

Theorem 3 : *Let B be a random variable whose domain is positive real numbers and let random variables* $V_1, V_2, \ldots, V_n, \ldots$ *be defined as* $V_1 = B/X_1$, $V_2 = V_1/X_2, \ldots$ $V_n = V_{n-1}/X_n$, *where* $X_1, X_2, \ldots, X_n$ *are independent and identically distributed random variables from* $U(0, 1)$. *Then*

$$\lim_{n \to \infty} \textit{Prob \{the m.s.d. of } V_n = A\} = \log_{10} \frac{A+1}{A}.$$

Proof : Consider the probability {the m.s.d. of $V_n = A \mid B = b$}. This conditional probability tends to $\log_{10} \frac{A+1}{A}$, which is independent of b. Hence the unconditional probability also tends to the same limit.

The two earlier theorems are particular cases of this theorem.

Acknowledgement

The author wishes to thank Dr. J. Roy and Dr. V. V. Menon and the referee for their helpful suggestions.

References

Adhikari, A. K. and Sarkar, B. P. (1968) : Distribution of most significant digit in certain functions whose arguments are random variables. *Sankhyā*, Series B, **30**, Parts 1 and 2, 47-58.

Benford, F. (1938) : Law of anomalous numbers. *Proc. Amer. Phil. Soc.* **78**, No. 4, 551-572.

Flehinger, B. J. (1966) : On the probability that a random integer has initial digit *A*. *Amer. Math. Monthly*, **73**.

Furry and Hurtwitz (1945) : Distribution of numbers and distribution of significant figures. *Nature*, **155**, 52-53.

Goudsmit and Furry (1944) : Significant figures of numbers in statistical tables. *Nature*, **154**, 800-801.

Pinkham, R. S. (1961) : On the distribution of first significant digits. *Ann. Math. Stat.*, **32**, 1223-1230.

Paper received : June, 1969.
Revised : October, 1969.

* The author is indebted to the referee for suggesting this.