Probability of Digits in a Decimal Expansion

Philip Pincencia
ppincencia@ucsd.edu

August 2024

I came across this paper that poses a problem regarding the probability of digits appearing in a position to the right of the decimal point.

Problem Statement

Given a well-behaved function $f : A \to B$, a digit x is uniformly chosen at random between $[f^{-1}(n), f^{-1}(n+1)]$, where $n \in \{\mathbb{Z} \cap A\}$. What is the probability that the digit in the r^{th}-position in the decimal expansion, d_r, is k?

This seems like a big problem that may be hard for us to wrap our head around what the final solution is going to look like. Let's do some examples of $f(x)$ and choose some values of n.

1 Examples

1.1 $f(x) = \sqrt{x}$

This means we choose a digit x uniformly at random from the interval $[n^2, (n+1)^2]$, where $n \in \mathbb{Z}^{\geq 0}$. An idea to tackle this problem is that if we get all non-overlapping sub-intervals such that $d_r = k$ between n^2 and $(n+1)^2$ with total length L, then $P(d_r = k) = \frac{L}{(n+1)^2 - n^2}$.

Let the set $C(d_r, k) = \{x \in [n^2, (n+1)^2] | \sqrt{x} = n.d_1 d_2 ... d_{r-1} \mathbf{k} d_{r+1} ...\}$. Then, $x \in c_{d_r,k}$ if and only if $\sqrt{x} = n.d_1 d_2 ... d_{r-1} \mathbf{k} d_{r+1} ...$. Now we want to bound $\sqrt{x}$ with something that's easier to work with. Notice that by truncating the decimals will always make the value smaller (like rounding down to say k decimals) and increasing the value of any decimal value will always make the value greater. Hence, this inequality

$$n.d_1 d_2 ... d_{r-1} k \leq \sqrt{x} \leq n.d_1 d_2 ... d_{r-1}(k+1)$$

holds. To make the decimal term with k the only decimal, we multiply all sides by 10^{r-1}, square all sides, and then isolate x

$$10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k}{10} \leq 10^{r-1}\sqrt{x} \leq 10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k+1}{10}$$

$$(10^{r-1}n + d_1 ... d_{r-1} + \frac{k}{10})^2 \leq 10^{2(r-1)} x \leq (10^{r-1}n + d_1 ... d_{r-1} + \frac{k+1}{10})^2$$

$$\left(\frac{10^{r-1}n + d_1 ... d_{r-1} + \frac{k}{10}}{10^{r-1}} \right)^2 \leq x \leq \left(\frac{10^{r-1}n + d_1 ... d_{r-1} + \frac{k+1}{10}}{10^{r-1}} \right)^2$$

$$\therefore x \in \left[\left(\frac{10^{r-1}n + d_1 ... d_{r-1} + \frac{k}{10}}{10^{r-1}} \right)^2, \left(\frac{10^{r-1}n + d_1 ... d_{r-1} + \frac{k+1}{10}}{10^{r-1}} \right)^2 \right]$$

Notice that the value $d_1 ... d_{r-1}$ can range from 0 (when all the $d_i = 0$) to $10^{r-1} - \mathbf{1}$ (when all the $d_i = 9$), and so we can let $q = d_1 ... d_{r-1} \in \{0, 1, ..., 10^{r-1} - 1\}$. And so
Now, we can use the formula for $P(d_r = k)$ immediately. Since there are 10^{r-1} values q can take, we have to sum it all up:

$$P(d_r = k) = \frac{1}{(n+1)^2 - n^2} \sum_{q=0}^{10^{r-1}-1} \frac{(10^{r-1}n + q + \frac{k+1}{10})^2 - (10^{r-1}n + q + \frac{k}{10})^2}{10^{2(r-1)}}$$

Using the difference of two squares and algebraic simplification , we get

$$\begin{aligned} P(d_r = k) = \frac{1}{2n+1} & \sum_{q=0}^{10^{r-1}-1} \frac{(2n \cdot 10^{r-1} + 2q + \frac{2k+1}{10})\frac{1}{10}}{10^{2(r-1)}} \\ &= \frac{10^{r-1}(2n \cdot 10^{r-1} + \frac{2k+1}{10}) + \sum_{q=0}^{10^{r-1}-1} 2q}{(2n+1)10^{2r-1}} \\ &= \frac{10^{r-1}(2n \cdot 10^{r-1} + \frac{2k+1}{10}) + (10^{r-1} - 1)10^{r-1}}{(2n+1)10^{2r-1}} \\ &= \frac{2n \cdot 10^{r-1} + \frac{2k+1}{10} + 10^{r-1} - 1}{(2n+1)10^r} = \frac{(2n+1) \cdot 10^r + 2k - 9}{(2n+1)10^{r+1}} \end{aligned}$$

, where $n \geq 0, k \geq 1, r \geq 1$.

Exercise for the readers:

Let $X \sim \text{Unif}[n^2, (n+1)^2]$. Given that $n = 3$, what is the most likely digit for the 4th digit of decimal expansion of $\sqrt{X}$?

1.2 $f(x) = \ln(x)$

This means we choose a digit x uniformly at random from the interval $[e^n, e^{n+1}]$, where $n \in \mathbb{Z}$. Again, We want to calculate the probability that the r^{th} digit in the decimal expansion of $\ln(x)$ is k in a similar way.

Define the set $C(d_r, k) = \{x \in [e^n, e^{n+1}] \mid \ln(x) = n.d_1 d_2 \ldots d_{r-1}\mathbf{k}d_{r+1}\ldots\}$. This implies that $\ln(x)$ must satisfy:

$$n.d_1 d_2 \ldots d_{r-1}k \leq \ln(x) \leq n.d_1 d_2 \ldots d_{r-1}(k+1)$$

Multiplying through by 10^{r-1} gives:

$$10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k}{10} \leq 10^{r-1}\ln(x) \leq 10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k+1}{10}$$

Exponentiate both sides to convert the logarithmic inequalities back into terms of x:

$$e^{\dfrac{10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k}{10}}{10^{r-1}}} \leq x < e^{\dfrac{10^{r-1}n + \sum_{m=2}^{r} 10^{r-m} d_{m-1} + \frac{k+1}{10}}{10^{r-1}}}$$

Thus, the digit x belongs to the interval:

$$x \in \left[e^{\frac{10^{r-1}n + q + \frac{k}{10}}{10^{r-1}}}, e^{\frac{10^{r-1}n + q + \frac{k+1}{10}}{10^{r-1}}} \right]$$

where $q = d_1 d_2 \ldots d_{r-1}$.

Hence,

$$P(d_r = k) = \frac{1}{e^{n+1} - e^n} \sum_{q=0}^{10^{r-1}-1} \left(e^{\frac{10^{r-1}n + q + \frac{k+1}{10}}{10^{r-1}}} - e^{\frac{10^{r-1}n + q + \frac{k}{10}}{10^{r-1}}} \right)$$

Using the approximation for small differences, $e^a - e^b \approx (a-b) \cdot e^b$, we can simplify the expression:

$$\begin{aligned} P(d_r = k) &\approx \frac{1}{e^{n+1} - e^n} \sum_{q=0}^{10^{r-1}-1} \left(\frac{k+1-k}{10^r} \cdot e^{\frac{10^{r-1}n + q}{10^{r-1}}} \right) \\ &\approx \frac{1}{e^{n+1} - e^n} \cdot \frac{1}{10^r} \sum_{q=0}^{10^{r-1}-1} e^{n + \frac{q}{10^{r-1}}} \\ &\approx \frac{1}{10^r(e-1)} \sum_{q=0}^{10^{r-1}-1} e^{\frac{q}{10^{r-1}}} \end{aligned}$$

Exercise for the readers:

Same as the previous exercise but now using $\ln(X)$ instead of $\sqrt{X}$ and $X \sim \text{Unif}[e^n, e^{n+1}]$

1.3 $f(x) = \cos^{-1}(x)$ TODO

This means we choose a digit x uniformly at random from the interval $[\cos(n), \cos(n+1)]$, where $n \in \mathbb{Z}$.