Report of CATAM 1.1 Random Binary Expansions

January 24, 2022

1 Monte Carlo Simulation

We can approximate F by randomly generating an n-digit sequence $(U_1, ..., U_n)$ using `numpy.random.choice`, computing $X^n = \sum_{i=1}^{n} \frac{U_i}{2^i}$. We loop this N times and implement a counter that increments if $X^n \leq x$ for our given x. Then we can plot the following:

$$\hat{F}(x) = \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}[X_j^n \leq x] \tag{1}$$

We choose to increment at $\Delta x = \frac{1}{216}$ as this will give us numbers with non-terminating expansions, such as $x = \frac{1}{3}$, and finite expansions, such as $x = \frac{1}{2}$. We choose $N = 1500$ to have low random error with reasonable computational time, and it clearly demonstrates that the methods agrees with each other in trend.

I have performed the simulation with larger values of $N = 5000, 10000$, but there was still observable noise (deviation from analytical solution) and not significant improvement in quality of result. In figure 1 below, the simulated results (red points) is plotted with the analytical solution for $p = \frac{2}{3}$ overlaid (blue line).

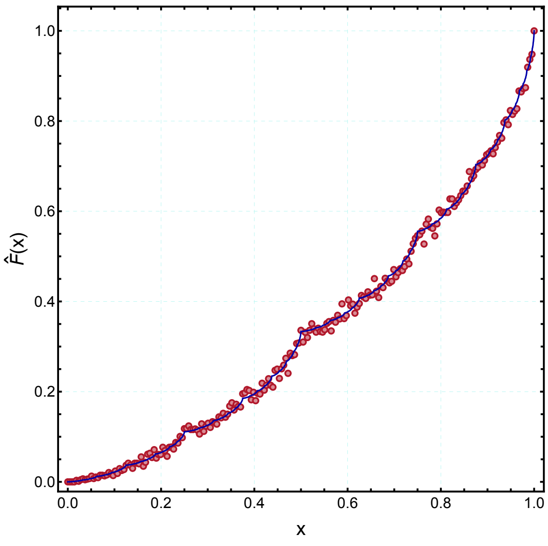

Figure 1: Monte Carlo Simulation: $p = \frac{2}{3}$ with Analytical Solution Overlaid

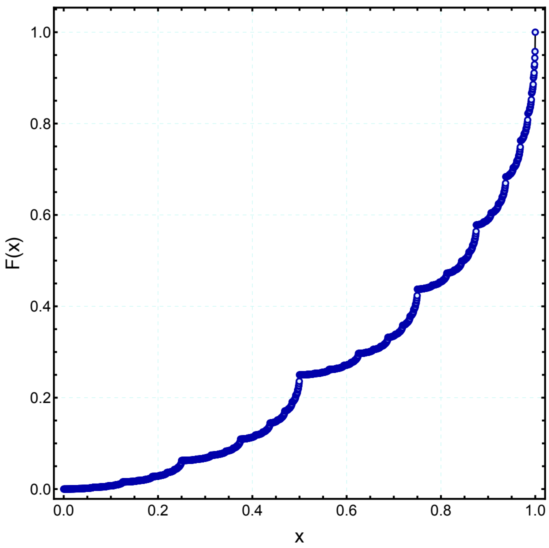

Figure 2: Analytical Solution: $p = \frac{3}{4}$

2 Formula for F

For a number $x = \sum_{i=1}^{n} \frac{X_i}{2^i}$, if the number $f(U) \leq x$, then let k be the first digit where the binary expansion differs. In order for them to differ, must have $U_k = 0, x_k = 1$. Before the k-th digit, all must be identical. Afterwards, they can take any value (and therefore does not contribute to the probability).

$$F(x) = \sum_{k=1}^{n} \mathbb{P}(U_i = x_i \ \forall \ 1 \leq i < k) \cdot \mathbb{P}(U_k = 0, x_k = 1) \tag{2}$$

The respective probability can be found by the term-by-term product of the probability reporting function $g(x)$:

$$g(x) = \begin{cases} p & x = 1 \\ q & x = 0 \end{cases} \tag{3}$$

$$F(x) = \sum_{k=1}^{n} \prod_{i}^{k-1} g(x_i) \cdot q \cdot \mathbb{I}(x_k = 1) \tag{4}$$

3 Comparison of Plot and Complexity

The plots demonstrate that the Monte Carlo simulation produces results in strong agreement with the analytical solution, albeit with some noise. In particular, when p is increased from $\frac{2}{3}$ to $\frac{3}{4}$, we observe that the cusps have a lower y-value overall. The cusps always occur at the same place: markedly so at $x = \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \ldots$. In fact they occur at all values which have finite binary expansion.

The complexity of the Monte Carlo Simulation $O(nN)$. We have a loop with N iterations, each generating a sequence of length n and checking its value (n addition operations).

The complexity of the analytical solution can be considered as follows: first we needed to get the binary expansion of x, which takes n subtraction operations. Then a loop with n iterations which checks the digits and multiplies the probabilities together. Overall it's $O(n)$. From some testing, the linear trend in computational time can only be seen when $n \approx 10$ (it takes a constant amount of time to import packages etc). Note that for actually plotting the graph, because there are 2^n points to compute, the time complexity is $O(n2^n)$.

4 Continuity of F

Claim: $F(x)$ is continuous at $x = c$, where c has a finite binary expansion

Proof: Write $c = \sum_{i=1}^{n} \frac{c_i}{2^n}$. For $c \in (0,1)$, $c_n = 1$ is the last non-zero digit. For $x \le 0$ and $x \ge 1$, $F(x) = 0$ and $F(x) = 1$ respectively, they are constant functions and therefore continuous on $x < 0$ and $x > 1$. From here on we consider $c \in [0,1]$.

Note that $F(x) = \mathbb{P}(X \le x)$ is an increasing function, and $F(x) - F(c)$ is precisely the probability that a generated number $f(U)$ lies in (x, c).

Case 1: $x > c$

If $c = 1$, $F(x) = F(c) = 1$. For $c < 1$, consider $\delta_1 < \min\left(\frac{1}{2^{n+m}}, 1 - c\right)$. When $c < x < c + \delta_1 < 1$, $F(c) \le F(x) \le F(c + \delta_1) \le 1$. For $c + \delta_1 < 1$ the first n-digits of its binary expansion are identical to that of c, then followed by $m - 1$ "0", and then a "1".

$$c + \delta_1 = 0.\ c_1\ c_2\ ...\ c_n\ 0\ 0\ ...\ 0\ 1$$

Then

$$\begin{aligned} F(x) - F(c) &= \mathbb{P}(U_i = c_i\ \forall\ 1 \le i \le n) \cdot \mathbb{P}(U_k = 0\ \forall n + 1 \le k \le n + m) \\ &\le \max(p,q)^n \cdot q^m \end{aligned} \tag{5}$$

Hence can choose $m \in \mathbb{N}$ such that $\max(p,q)^n \cdot q^m < \epsilon$.

That means that for $0 < x - c < \delta_1$, $0 \le F(x) - F(c) < \epsilon$.

Case 2: $x < c$

If $c = 0$, then $F(x) = F(c) = 0$. For $0 < c \le 1$, consider $\delta_2 = \min(\frac{1}{2^{n+r}}, c)$. When $0 \le c - \delta_2 < x < c$, $0 \le F(c - \delta_2) \le F(x) \le F(c)$. For $c - \delta_2$, the first $n - 1$ -digits of its binary expansion are identical to that of c, then followed by a "0" and r "1".

$$x = 0.\ c_1\ c_2\ ...\ c_{n-1}\ 0\ 1\ 1\ ...\ 1\ 1$$

$$\begin{aligned} F(c) - F(x) &= \mathbb{P}(U_i = c_i\ \forall\ 1 \le i \le n - 1) \cdot \mathbb{P}(U_n = 0) \cdot \mathbb{P}(U_i = 0\ \forall n + 1 \le i \le n + r) \\ &\le \max(p,q)^n \cdot q \cdot p^r \end{aligned} \tag{6}$$

Hence can choose $r \in \mathbb{N}$ such that $\max(p,q)^n \cdot q \cdot p^r < \epsilon$.

That means that for $-\delta_2 < x - c < 0$, $-\epsilon < F(x) - F(c) \le 0$.

Take $\delta = \min(\delta_1, \delta_2)$, then given $\epsilon > 0$, if $|x - c| < \delta$, then $|F(x) - F(c)| < \epsilon$.

Claim: $F(x)$ is continuous when x has a non-terminating binary expansion.

Proof: Previously we have shown that $F(x)$ is a constant function and hence continuous on $x < 0$ and $x > 1$. Now consider $x \in (0,1)$. If x has a non-terminating binary expansion (we do not consider representations that have consecutive "1" after some point, since they can be equivalently written in a terminating form), then given $n \in \mathbb{N}$, there exist $m \geq n$ such that $x_m = 1, x_{m+1} = 0$.

Then consider the numbers

$$a = 0.x_1\ x_2\ ...\ x_m\ 0\ 0\ ...$$
$$b = 0.x_1\ x_2\ ...\ x_m\ 1\ 1\ ...$$

Because $a < x < b$, then $F(a) \leq F(x) \leq F(b)$ and hence $|F(y) - F(x)| \leq F(b) - F(a)$ $\forall y \in (a, b)$.

$$\begin{aligned} F(b) - F(a) &= \prod_{i=1}^{m} \mathbb{P}(u_i = x_i) \\ &\leq \max(p, q)^m \end{aligned} \tag{7}$$

So can pick $n \in \mathbb{N}$ s.t. $\max(p,q)^m \leq \max(p,q)^n < \epsilon$. Then pick $\delta = \min(b - x, x - a)$ so that $\forall |y - x| < \delta, |F(y) - F(x)| < \epsilon$.

5 $\frac{F(c+\delta)-F(c)}{\delta}$ for $c + \delta$ with Finite Binary Expansion

We plot $\frac{F(c+\delta)-F(c)}{\delta}$ for $c + \delta$ for $c = \frac{9}{16}$ and $\delta = \frac{1}{2^{12}} \times m$, $m = -32, -31, ..., 0, 1, ..., 32$. Hence the range investigated is $\left[\frac{9}{16} - \frac{1}{128}, \frac{9}{16} + \frac{1}{128}\right]$, sufficiently small for limit behaviour.

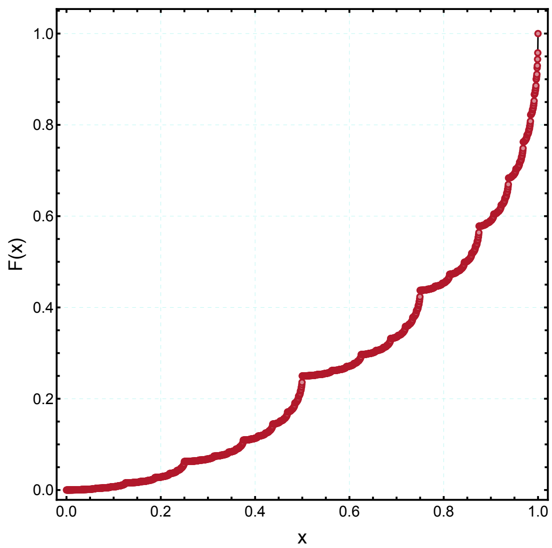

Figure 3: $F(x)$ when $p = \frac{3}{4}$

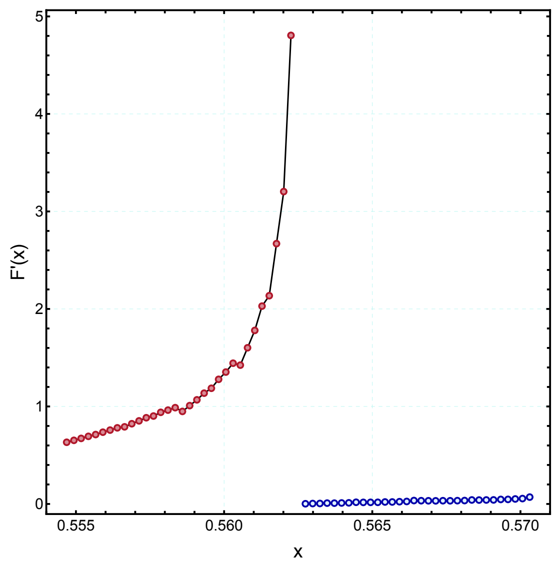

Figure 4: $\frac{F(x)-F(c)}{x-c}$ when $c = \frac{9}{16}$, $p = \frac{3}{4}$

It is apparent that when $\delta < 0$, the graph diverges. When $\delta > 0$, the graph tends to 0. The scale chosen demonstrate the diverging trend and that the magnitude for points on the left is much larger than the magnitude for points on the right. There was no need for greater vertical plot range as the trend is apparent. These plots suggest that F is right-differentiable but not left-differentiable at $c = \frac{9}{16}$.

6 When is F differentiable?

Claim: For $c \in (0,1)$ with a finite binary expansion,

When $p > \frac{1}{2}$, $F(x)$ is right-differentiable but not left-differentiable at $x = c$. F is also differentiable at $x = 0$.

When $p = \frac{1}{2}$, $F(x)$ is differentiable and $F'(x) = 1$.

When $p < \frac{1}{2}$, $F(x)$ is left-differentiable but not right-differentiable at $x = c$. F is also differentiable at $x = 1$

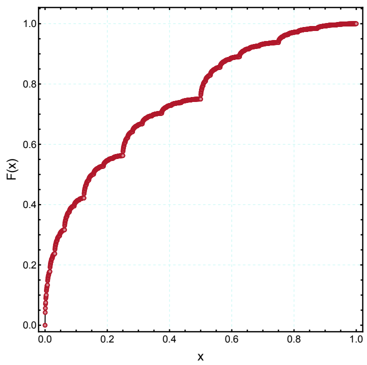

Figure 5: $F(x)$ when $p = \frac{1}{4}$

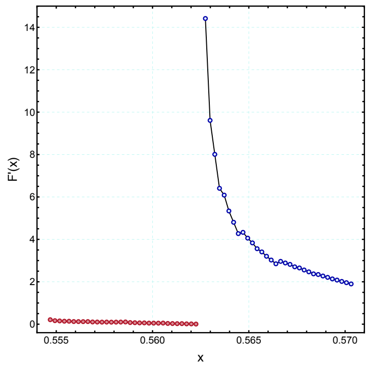

Figure 6: $F'\left(\frac{9}{16}\right)$ when $p = \frac{1}{4}$

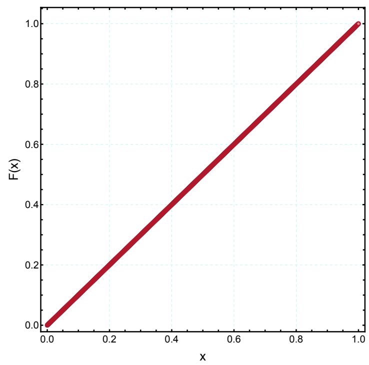

Figure 7: $F(x)$ when $p = \frac{1}{2}$

Proof:

When $p = \frac{1}{2}$, the probability of generating any number is uniform in $[0, 1]$ as it does not depend on the digits in its binary representation. Hence $F(x) = \mathbb{P}(f(U) \leq x) = x$, $F'(x) = 1$ for $x \in (0, 1)$.

When $p \neq \frac{1}{2}$:

Right Differentiable

Consider $0 < \delta < \frac{1}{2^{n+1}}$ that has a finite binary expansion. That means the digits of $c + \delta$ are the digits of c, followed by some zeros, then the digits of δ. Here we denote δ_m as the first non-zero digit of δ.

$$c + \delta = 0.\ c_1\ c_2\ \ldots\ c_n\ 0\ 0\ \ldots\ 0\ \delta_m\ \delta_{m+1}\ \ldots\ \delta_{m+r}$$

Then

$$\begin{aligned} 0 \leq F(c+\delta) - F(c) &\leq F\left(c + \frac{1}{2^{m-1}}\right) - F(c) \\ &= \left(\prod_{i=1}^{n} \mathbb{P}(U_i = c_i)\right) \cdot q^{m-n-1} \end{aligned} \tag{8}$$

Then

$$\begin{aligned} 0 \leq \frac{F(c+\delta) - F(c)}{\delta} &\leq \frac{\left(\prod_{i=1}^{n} \mathbb{P}(U_i = c_i)\right) \cdot q^{m-n-1}}{\frac{1}{2^m}} \\ &= \prod_{i=1}^{n} \mathbb{P}(U_i = c_i) \cdot q^{-n-1} \cdot (2q)^m \end{aligned} \tag{9}$$

If $p > \frac{1}{2}, q < \frac{1}{2}, \delta \to 0$, then $m \to \infty$ and $(2q)^m \to 0$. Hence this limit exists and

$$\lim_{\delta \downarrow 0} \frac{F(c+\delta) - F(c)}{\delta} = 0 \tag{10}$$

If $p < \frac{1}{2}$, this limit does not exist and $F(x)$ is not right-differentiable.

Left Differentiable

Consider $0 < \delta < \frac{1}{2^{n+1}}$ that has a finite binary expansion. Let its first non-zero digit be in the m-th place (i.e. $\frac{1}{2^m} \leq \delta < \frac{1}{2^{m-1}}$)

$$c - \delta = 0.\ c_1\ c_2\ ...\ c_{n-1}\ 0\ 1\ 1\ ...\ 1\ \mu_m\ \mu_{m+1}\ ...\ \mu_{m+r}$$

Then

$$\begin{aligned} 0 \geq F(c-\delta) - F(c) &\geq F\left(c - \frac{1}{2^{m-1}}\right) - F(c) \\ &= -\left(\prod_{i=1}^{n-1} \mathbb{P}(U_i = c_i)\right) \cdot q \cdot p^{m-n-1} \end{aligned} \tag{11}$$

Then

$$\begin{aligned} 0 \geq \frac{F(c-\delta) - F(c)}{-\delta} &\geq \frac{\left(\prod_{i=1}^{n-1} \mathbb{P}(U_i = c_i)\right) \cdot q \cdot p^{m-n-1}}{\frac{1}{2^m}} \\ &= \prod_{i=1}^{n-1} \mathbb{P}(U_i = c_i) \cdot q \cdot p^{-n-1} \cdot (2p)^m \end{aligned} \tag{12}$$

If $p < \frac{1}{2}, \delta \to 0$, then $m \to \infty$ and $(2p)^m \to 0$. Hence this limit exists and

$$\lim_{\delta \uparrow 0} \frac{F(c+\delta) - F(c)}{\delta} = \lim_{\delta \downarrow 0} \frac{F(c-\delta) - F(c)}{-\delta} = 0 \tag{13}$$

If $p > \frac{1}{2}$, this limit does not exist and $F(x)$ is not right-differentiable.