APMA1690: Homework # 3 (Due by 11pm Oct 5)

1 Review

- (**Multiplicative Congruential Generator**)
  For "properly chosen" positive integers m and a (e.g., $m = 2^{31} - 1$ and $a = 7^5$), we have the multiplicative congruential generator (MCG) presented by the following algorithm

---

Algorithm 1 : Multiplicative Congruential Generator

Input: (i) positive integers m and a; (ii) a seed $s \in \{1, 2, \ldots, m-1\}$; (iii) sample size n.
Output: pseudo-random numbers $g(1), g(2), \ldots, g(n)$. (These numbers look like iid from $\text{Unif}(\{1, 2, \ldots, m-1\})$.)

```
1: Initialization: g(1) ← s.
2: for all i = 1, 2, ..., n − 1, do
3:     g(i + 1) ← (a · g(i) mod m) = the remainder of a·g(i)/m.
4: end for
```

The "remainder" referred to in the algorithm above is the remainder in integer division.

- (**Fundamental theorem for the "inverse CDF method"**)
  Let $F(t)$ be a CDF of interest. We define the function $G(u)$ on the open interval $(0, 1)$ by

$$G(u) = \inf \{t \in \mathbb{R} : F(t) \geq u\}, \quad \text{for all } u \in (0,1), \tag{1}$$

where "inf" denotes the infimum operation (you may view it as "min" for simplicity). The function G in Eq. (1) is the "generalized inversion" of F. Let U be a random variable defined on the probability space $(\Omega, \mathbb{P})$ and following the continuous uniform distribution on $(0, 1)$, i.e., $U \sim \text{Unif}(0, 1)$, and we define a new random variable X by

$$X(\omega) = G(U(\omega)), \quad \text{for all } \omega \in \Omega.$$

Then, the CDF of X is the CDF $F(t)$ of interest.
Remarks: (i) For any given real number $0 < u < 1$, the notation "$\{t \in \mathbb{R} : F(t) \geq u\}$" denotes the collection of the real numbers t such that $F(t) \geq u$.
(ii) "$\inf \{t \in \mathbb{R} : F(t) \geq u\}$" denotes the smallest number in the collection $\{t \in \mathbb{R} : F(t) \geq u\}$.
(iii) If the inverse F^{-1} of F exists, then $G = F^{-1}$.

- Algorithm 2 algorithm provides the procedures for implementing the inverse CDF method.

Algorithm 2 : Inverse CDF Method

Input: (i) The CDF F of interest; (ii) sample size n.
Output: A sequence of iid (pseudo) random numbers $x_1, \ldots, x_n$ following the distribution F.

Generate (pseudo) iid random variables $u_1, u_2, \ldots, u_n$ from $\mathrm{Unif}(0, 1)$.
for all $i = 1, \ldots, n$, **do**
  Compute $x_i \leftarrow G(u_i) = \inf\{t \in \mathbb{R} : F(t) \geq u_i\}$.
end for

2 Problem Set

1. This question helps you better understand the MCG. Let g be the MCG in Algorithm 1.

   (a) (1 point) Write the computer code for the MCG as described in Algorithm 1 using your preferred programming language. Please include your code in your submission.

```
def MCG(s, n):
    m = 2**(31) - 1
    a = 7**5
    g = [s]

    for i in range(1, n):
        g.append((a * g[i - 1]) % m)

    return g
```

   (b) (1 point) Let $m = 2^{31} - 1$ and $a = 7^5$. Initialize the seed by $g(1) \leftarrow 1690$ and generate $g(1), g(2), \ldots, g(10)$ using your code. Show all the ten numbers $g(1), g(2), \ldots, g(10)$ and the code for generating these numbers.

n	$g(n)$
1	1690
2	28403830
3	641801176
4	2089489798
5	254955595
6	808809400
7	88100290
8	1085341247
9	604240711
10	23463114

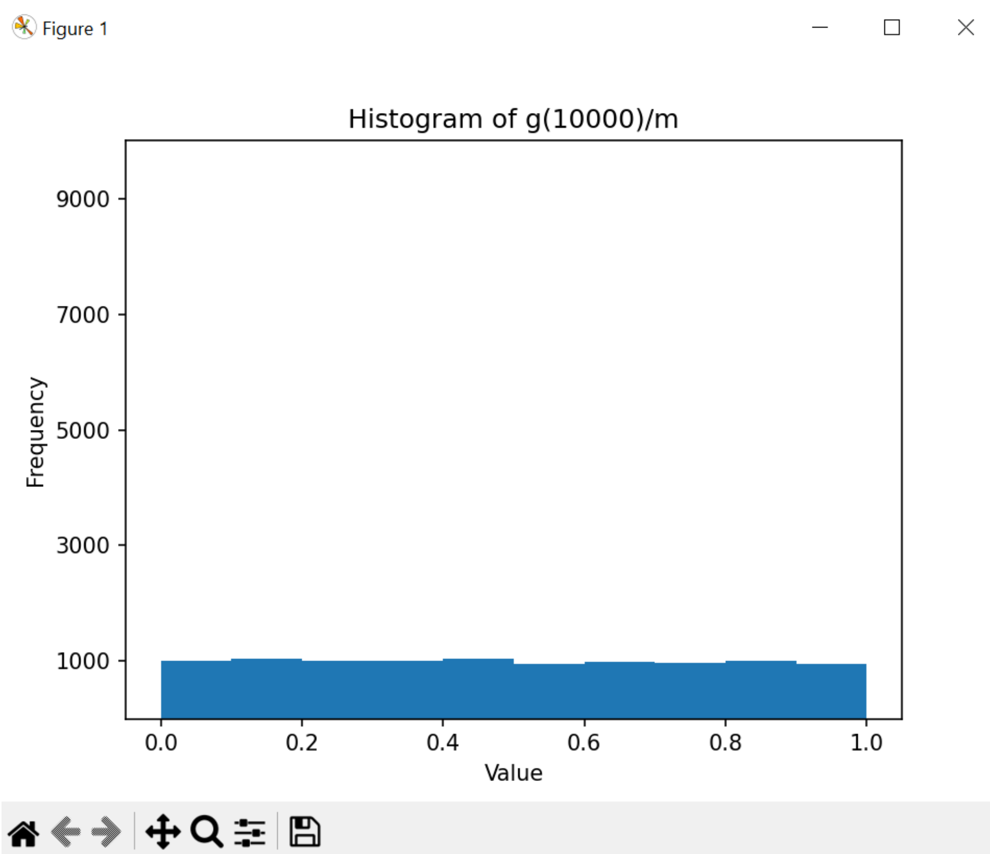
```
def Q1B():
    return MCG(s=1690, n=10)
```

   (c) (1 point) Let $m = 2^{31} - 1$ and $a = 7^5$. Initialize the seed by $g(1) \leftarrow 1690$ and generate $g(1), g(2), \ldots, g(10000)$ using your code. Plot and show the histogram of the following

10000 values

$$\left\{ \frac{g(1)}{m}, \frac{g(2)}{m}, \ldots, \frac{g(10000)}{m} \right\}.$$

Show the code for generating this histogram.




```python
def Q1F(plot=True):
    n = 10000
    m = 2**(31) - 1
    s = 1690

    y = list(map(lambda x: math.log(1/(1 - (x/m))), MCG(s=s, n=n)))

    if plot:
        plt.title("Random from Exp(1)")
        plt.ylabel("Frequency")
        plt.xlabel("Value")
        plt.ylim(0, 10000)
        plt.yticks(range(1000, 10000, 2000), range(1000, 10000, 2000))
        plt.hist(y)
        plt.show()

    return y
```

(d) (1 point) Please heuristically (rather than mathematically/rigorously) explain the relationship between the histogram you obtained in the preceding question and Unif$(0, 1)$.

The sequence of random numbers $\left\{\frac{g(1)}{m}, \ldots, \frac{g(n)}{m}\right\}$ "look like" random numbers taken iid from Unif$(0, 1)$.

(e) (0.5 points) Let $F(t) = (1 - e^{-t}) \cdot \mathbb{1}(t > 0)$, which is the CDF of the exponential distribution Exp(1). Compute an explicit express of the $G(u)$ defined in Eq. (1) for all $0 < u < 1$.

By the Fundamental Theorem of the Inverse CDF, we are looking for an explicit expression for

$$G(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\} \quad \forall u \in (0, 1)$$

Heuristically, we ask "what is the smallest t for which $F(t)$ is greater than or equal to u, for all $u \in (0, 1)$?" First, we observe

$$F(t) = (1 - e^{-t}) \cdot \mathbb{1}(t > 0) = \begin{cases} 1 - e^{-t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

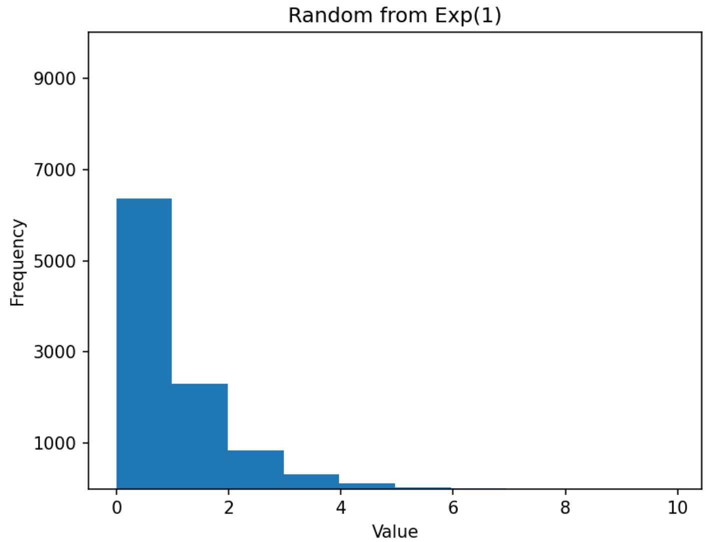
Immediately, we see that $G(u)$ is not defined for $t \leq 0$ and for $t > 0$:

$$\begin{aligned} G(u) &= \inf\{t \in \mathbb{R} : 1 - e^{-t} \geq u\} \\ &= \inf\{t \in \mathbb{R} : 1 - u \geq e^{-t}\} \\ &= \inf\{t \in \mathbb{R} : \log(1 - u) \geq -t\} \\ &= \inf\{t \in \mathbb{R} : -\log(1 - u) \leq t\} \\ &= \inf\{t \in \mathbb{R} : \log(\frac{1}{1 - u}) \leq t\} \\ &= \boxed{\log(\frac{1}{1 - u})} \end{aligned}$$

(f) (0.5 points) Using the values $g(1), g(2), \ldots, g(10000)$ generated in the preceding question, plot and show the histogram of the following 10000 values

$$\left\{\log\left(\frac{1}{1 - \frac{g(1)}{m}}\right), \log\left(\frac{1}{1 - \frac{g(2)}{m}}\right), \ldots, \log\left(\frac{1}{1 - \frac{g(10000)}{m}}\right)\right\},$$

where "log" denotes the natural logarithm and $m = 2^{31} - 1$. Please heuristically (rather than mathematically/rigorously) explain the relationship between the histogram you obtained here and the probability density function of the exponential distribution Exp(1).




```python
def Q1F(plot=True):
    n = 10000
    m = 2**(31) - 1
    s = 1690

    y = list(map(lambda x: math.log(1/(1 - (x/m))), MCG(s=s, n=n)))

    if plot:
        plt.title("Random from Exp(1)")
        plt.ylabel("Frequency")
        plt.xlabel("Value")
        plt.ylim(0, 10000)
        plt.yticks(range(1000, 10000, 2000), range(1000, 10000, 2000))
        plt.hist(y)
        plt.show()

    return y
```

By the inverse CDF method, this sequence of values are random numbers which “look like” random numbers generated iid from the distribution Exp(1).

2. Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$, and X satisfies

$$\mathbb{P}\{\omega \in \Omega : X(\omega) = i\} = \frac{1}{n}, \quad \text{for all } i = 1, 2, \ldots, n.$$

where n is a given positive integer. Define a new random variable Y by

$$Y(\omega) = \frac{X(\omega)}{n}, \quad \text{for all } \omega \in \Omega.$$

(a) (0.5 point) Show that the CDF of Y is the following

$$\begin{aligned} F_n(t) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\frac{i}{n} \le t\}} \\ &= \frac{\text{the number of integers } i \text{ such that } \frac{i}{n} \le t}{n}, \end{aligned} \tag{2}$$

for all $t \in \mathbb{R}$.

We first note that from the definition of a discrete CDF and $\mathbb{P}(X = i) = \frac{1}{n}$, the CDG od X is

$$F_X(t) = \sum_{i=1}^{n} \frac{1}{n} \cdot \mathbb{1}_{(i \le t)}$$

Then, because $Y(\omega) = \frac{X(\omega)}{n}$,

$$F_Y(t) = \mathbb{P}(Y \le t) = \mathbb{P}(\frac{X}{n} \le t) = \sum_{i=1}^{n} p_i \cdot \mathbb{1}_{(y_i \le t)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(\frac{i}{n} \le t)} \quad \blacksquare$$

(b) (1 point) Let $F_n(t)$ be the CDF defined in Eq (2). Use the definition of Riemann integrals to prove the following

$$\begin{aligned} \lim_{n\to\infty} F_n(t) &= \text{ the CDF of } \mathrm{Unif}(0,1) \\ &= \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } 0 \le t < 1 \\ 1 & \text{if } 1 \le t. \end{cases} \end{aligned}$$

From Eq 2,

$$\lim_{n\to\infty} F_n(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\frac{i}{n} \le t\}}$$

Since $1 \le i \le n$, $\frac{1}{n} \le \frac{i}{n} \le 1$ so we know that $\mathbb{1}_{(x \le t)} = 0$ for $t < 0$. Similarly, for $t \ge 1$, $\mathbb{1}_{(x \le 1)} = 1$ for all x.
But we notice that the definition of Riemann Integrals,

$$\int_0^1 H(x)\,dx = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(\frac{i}{n})$$

is quite similar to $\lim_{n\to\infty} F_n(t)$, especially since we only need to check $0 \le t < 1$. Letting $H(\frac{i}{n}) = \mathbb{1}_{\frac{i}{n}\le t}$, we have

$$\begin{aligned}
\lim_{n\to\infty} F_n(t) &= \int_0^1 \mathbb{1}_{(x\le t)}\, dx \\
&= \int_0^t 1\, dx + \int_t^1 0\, dx \\
&= t
\end{aligned}$$

All together,

$$\lim_{n\to\infty} F_n(t) = \begin{cases} 0 & t < 0 \\ t & 0 \le t < 1 \\ 1 & \le t \end{cases} \qquad \blacksquare$$

3. Let $F(t)$ denote the CDF of the Bernoulli($\frac{1}{3}$) distribution.

(a) (1 point) Compute an explicit express of the $G(u)$ defined in Eq. (1) for all $0 < u < 1$.

$$F(t) = \frac{2}{3} \cdot \mathbb{1}_{(t \geq 0)} + \frac{1}{3} \cdot \mathbb{1}_{(t \geq 1)}$$

So

$$\begin{aligned} G(u) &= \inf\{t \in \mathbb{R} : F(t) \geq u\} \\ &= \inf\{t \in \mathbb{R} : \frac{2}{3} \cdot \mathbb{1}_{(t \geq 0)} + \frac{1}{3} \cdot \mathbb{1}_{(t \geq 1)} \geq u\} \\ &= \boxed{\begin{cases} 0 & \text{if } 0 < u \leq \frac{2}{3} \\ 1 & \text{if } \frac{2}{3} < u < 1 \end{cases}} \end{aligned}$$

(b) (1 point) Explain why $G(0) = \inf\{t \in \mathbb{R} : F(t) \geq 0\}$ is ill-defined.
In Eq 2, u is defined on the *open* interval $(0, 1)$ so $u = 0$ is not in the domain of the inverse.

(c) (0.5 point) Let $g(1), g(2), \ldots, g(10)$ be the values you generated in Question 1 (b). Compute and show the following ten values

$$\left\{G\left(\frac{g(1)}{m}\right), G\left(\frac{g(2)}{m}\right), \ldots, G\left(\frac{g(10)}{m}\right)\right\},$$

where $m = 2^{31} - 1$ and G is the function in Question 3 (a).

$$\{g(i)\}_{i=1}^{10} = \{0, 0, 0, 1, 0, 0, 0, 0, 0, 0\}$$

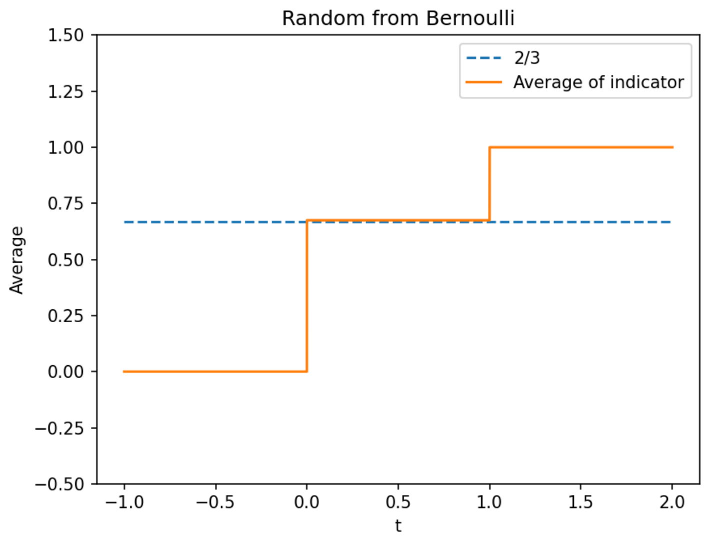
```
def Q3C():
    m = 2**(31) - 1
    g = Q1B()

    def G(n):
        if n < (2 / 3):
            return 0
        else:
            return 1

    return list(map(lambda x: G(x/m), g))
```

(d) (1 point) Let $g(1), g(2), \ldots, g(10000)$ be the values you generated in Question 1 (c). Denote

$$x_i = G\left(\frac{g(i)}{m}\right), \quad \text{for all } i = 1, 2, \ldots, 10000,$$

where $m = 2^{31} - 1$. Plot and show the graph of the following function of t

$$\frac{1}{10000} \sum_{i=1}^{10000} \mathbf{1}\{x_i \leq t\}.$$




```python
def Q3D(plot=True):
    m = 2**(31) - 1
    g = Q1C(False)

    def G(n):
        if n < (2 / 3):
            return 0
        else:
            return 1

    def ind_avg(x):
        y = []
        total = 0
        n = range(len(x))

        for index in n:
            total += x[index]
            y.append(total)
        y = list(map(lambda x: x/len(n), y))

        return y

    t = np.linspace(-1, 2, 10000)
    y = ind_avg(list(map(lambda x: G(x), g)))

    if plot:
        plt.title("Random from Bernoulli")
        plt.ylabel("Average")
        plt.xlabel("t")
        plt.ylim(0, 0.5)
        plt.plot(t, np.linspace(1/3, 1/3, 10000), linestyle='dashed', label='1/3')
        plt.plot(t, y, label='Average of indicator')

        plt.legend()
        plt.show()

    return y
```