Statistical Inference Project 1

Question 1(a)

- Aim:- Device an algorithm using inverse transformation technique to generate random number from the above probability density function.

According to inverse algorithm we need to find cumulative distribution function (CDF) for the given PDF.

$$F(x) = \int_0^x (2u+1)e^{-(u^2+u)}du$$

$$\text{using} \quad t = u^2 + u, \quad dt = 2udu + du = (2u+1)du$$

$$F(x) = \int_0^x e^{-t}dt$$

$$F(x) = [-e^t]_0^x$$

$$F(x) = 1 - e^{-(x^2+x)}$$

cumulative distribution function (CDF) for the given PDF:

$$F(x) = \begin{cases} 1 - e^{-(x^2+x)} & x > 0 \\ 0 & otherwise \end{cases}$$

Finding quasi-inverse of F:-

$$F(x) = 1 - e^{-(x^2+x)}$$

$$e^{-(x^2+x)} = 1 - F(x)$$

$$\text{taking log both sides}$$

$$-(x^2+x) = ln(1-F(x))$$

$$(x^2+x) + ln(1-F(x)) = 0$$

$$\text{solving for x}$$

$$x = \frac{-1 \pm \sqrt{1 - 4ln(1-F(x))}}{2}$$

$$x = \frac{-1}{2} \pm \sqrt{\frac{1}{4} - ln(1-F(x))}$$

$$\text{as x} > 0$$

$$x = \frac{-1}{2} + \sqrt{\frac{1}{4} - ln(1-F(x))}$$

Now according to the inverse algorithm, making

$$F^{-1}(U) = \frac{-1}{2} + \sqrt{\frac{1}{4} - ln(1-U)}$$

replacing 1-U with U

$$F^{-1}(U) = \frac{-1}{2} + \sqrt{\frac{1}{4} - ln(U)}$$

Algorithm is given by:

1. Generate U from U(0, 1) distribution.

2. $X = \frac{-1}{2} + \sqrt{\frac{1}{4} - ln(U)}$

3. Return X

Question 1(b)

- Aim:- Obtain an estimate of the expectation corresponding to the above probability density function using Monte Carlo technique. Mention each step in the report including the value of seed and the number of random numbers used in this calculation.

In a simple Monte Carlo problem, we express the quantity we want to know as the expected value of a random variable Y, such as `μ = E(Y)`. Generate values `Y1, . . . , Yn` independently from the distribution of Y.

Take their average

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}Y_i$$

as an estimate of μ.

In this case we take random sample of size `10,0000` using the seed number `42`.
which can be done by generating 10,000 random numbers from U(0, 1) and the using X from part(a) converting it into given pdf distribtution.
Then taking average of these random sample (X) we find estimator for expected value.
which comes as `0.5456065888744848`.
To verify this estimator we use technology to integrate xf(x) which is approximately `μ = 0.5456`.

Question 1(c)

Again using the same seed that we used in part(b) `42`.
This time we are generating 10,000 with same procedure.
Using the graph we can see that histogram of generated random number is following the same distribution of the given pdf. which is obvious and also this verifing our algorithm.



Histogram with Probability Density Function
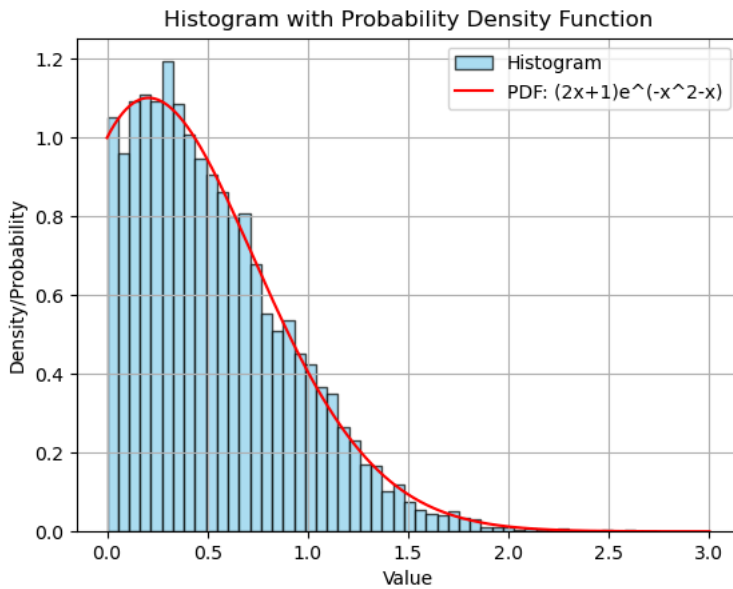
Question 1(d)

For the given random sample $(x_1, x_2, ...x_n)$ expirical function is defined as the discrete distribution that assigns to every value X_k the probability $\frac{1}{n}$.
Hence Empirical function:

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_k < x \le x_{k+1}, \quad 1 \le k \le n-1 \\ 1 & x > x_n \end{cases}$$

And given CDF:

$$F(x) = \begin{cases} \left(1 - e^{-\frac{x}{\theta}}\right)^2 & x > 0 \\ 0 & x \le 0 \end{cases}$$

Ploting these functions on same graph:

Empirical Distribution Function and $1 - e^{-x^2 - x}$
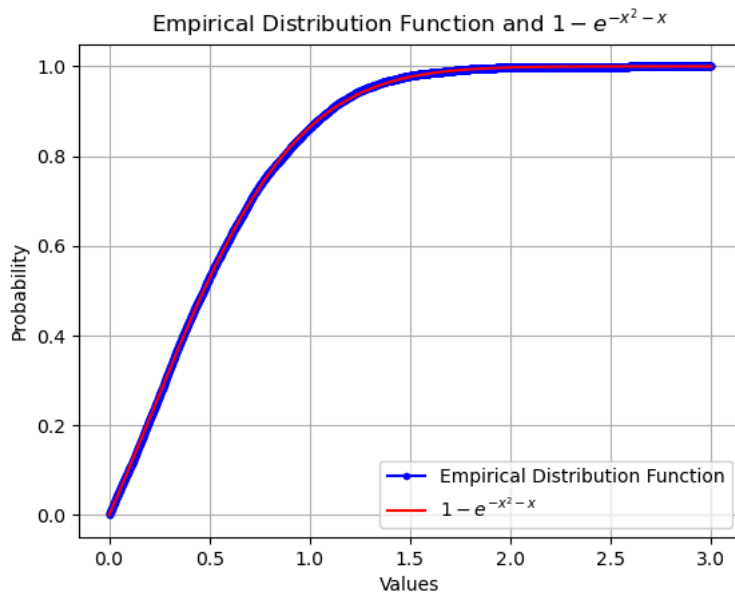
We can clearly see that both of these functions have same curve, which can be seen as:

$$F_n(x) \to F(x), \quad when \quad x \to \infty$$

Question 2(a)

Given CDF:

$$F(x) = \begin{cases} \left(1 - e^{-\frac{x}{\theta}}\right)^2 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Evaluating Probability distribution function (PDF) by differentiating given CDF:

$$f(x|\theta) = \frac{d}{dx} F(x|\theta)$$

$$f(x|\theta) = \frac{d}{dx} \left(1 - e^{-\frac{x}{\theta}}\right)^2$$

$$f(x|\theta) = 2 \left(1 - e^{-\frac{x}{\theta}}\right) \frac{d}{dx} \left(1 - e^{-\frac{x}{\theta}}\right)$$

$$f(x|\theta) = 2 \left(1 - e^{-\frac{x}{\theta}}\right) \left(\frac{1}{\theta} e^{-\frac{x}{\theta}}\right)$$

$$f(x|\theta) = \frac{2}{\theta} \left(1 - e^{-\frac{x}{\theta}}\right) e^{-\frac{x}{\theta}}$$

Hence the probability density function is given by:

$$f(x|\theta) = \begin{cases} \frac{2}{\theta} \left(1 - e^{-\frac{x}{\theta}}\right) e^{-\frac{x}{\theta}} & x \geq 0 \\ 0 & otherwise \end{cases}$$

Now we can define the likelihood function as follows:

$$L(\theta) = \prod_{i=1}^{53} f(x_i|\theta)$$

$$L(\theta) = \prod_{i=1}^{53} \frac{2}{\theta} \left(1 - e^{-\frac{x_i}{\theta}}\right) e^{-\frac{x_i}{\theta}}$$

$$L(\theta) = \left(\frac{2}{\theta}\right)^{53} \prod_{i=1}^{53} \left(1 - e^{-\frac{x_i}{\theta}}\right) e^{-\frac{x_i}{\theta}}$$

The likelihood function is given by:

$$F(x) = \begin{cases} L(\theta) = \left(\frac{2}{\theta}\right)^{53} \prod_{i=1}^{53} \left(1 - e^{-\frac{x_i}{\theta}}\right) e^{-\frac{x_i}{\theta}} & x_i \geq 0 \\ 0 & otherwise \end{cases}$$

Taking log-likelihood to make it more suitable to find MLE:

$$ln(L(\theta)) = 53ln\left(\frac{2}{\theta}\right) + \sum_{i=1}^{53} ln\left(1 - e^{-\frac{x_i}{\theta}}\right) - \sum_{i=1}^{53} \frac{x_i}{\theta}$$

$$ln(L(\theta)) = 53ln(2) - 53ln(\theta) + \sum_{i=1}^{53} ln\left(1 - e^{-\frac{x_i}{\theta}}\right) - \sum_{i=1}^{53} \frac{x_i}{\theta} \tag{1}$$

To find the MLE, we have to maximize likelihood function which is equivalent to maximum log-likelihood function as log is increasing function. Now, to maximum log-likelihood we use derivative test:

First we evaluate derivative of log-likelihood function to find critical points making it equal to 0:

$$\frac{d}{d\theta}ln(L(\theta)) = 0$$

$$\frac{d}{d\theta}\left(53ln(2) - 53ln(\theta) + \sum_{i=1}^{53} ln\left(1 - e^{-\frac{x_i}{\theta}}\right) - \sum_{i=1}^{53} \frac{x_i}{\theta}\right) = 0$$

$$-\frac{53}{\theta} - \sum_{i=1}^{53} \frac{e^{-\frac{x_i}{\theta}}x_i}{\theta^2(1 - e^{-\frac{x_i}{\theta}})} + \sum_{i=1}^{53} \frac{x_i}{\theta^2} = 0$$

here we can clearly see that the above equation is not solvable analytically, so we will use numerical methods to find the MLE. We will use the gradient descent method to find the MLE. The gradient descent method is an iterative method to find the minimum of a function. The algorithm is as follows:

1. Initialize the value of θ
2. Calculate the gradient of the function at the current value of θ
3. Update the value of θ using the formula:

$$\theta_{new} = \theta_{old} - \alpha\frac{d}{d\theta}ln(L(\theta))$$

where α is the learning rate.

4. Repeat steps 2 and 3 until the value of θ converges.

We will use the following values for the parameters:

1. Initial value of θ: 1
2. Learning rate: 0.001
   We will use the following stopping criteria:
3. 1000 iterations
   since our problem is of maximization, we can convert it into a minimization problem by multiplying it by -1. we also have one constant term in the equation (1) i.e 53ln2. which can be discarded because it won't affect the optimum solution. So our final equation for minimization is:

$$-ln(L(\theta)) = -\sum_{i=1}^{53} ln\left(1 - e^{-\frac{x_i}{\theta}}\right) + \sum_{i=1}^{53} \frac{x_i}{\theta} + 53ln(\theta) \tag{2}$$

$$\frac{d}{d\theta}ln(-L(\theta)) = \frac{1}{\theta^2}\left(53\theta + \sum_{i=1}^{53} \frac{e^{-\frac{x_i}{\theta}}x_i}{(1 - e^{-\frac{x_i}{\theta}})} - \sum_{i=1}^{53} x_i\right)$$

using the above equation with specified parameters, we have found that the MLE is converging to 0.5 as we can see in the below graph:
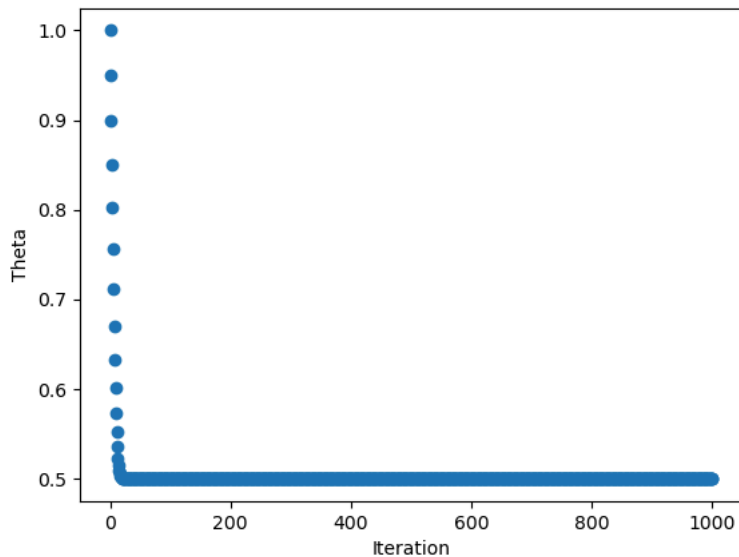
Question 2(b)

According to the given question we want to estimate the bias and mean squared error of the MLE of θ = 0.5, 1,2 using Monte-Carlo method. The bias and mean squared error are defined as follows:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$
$$\text{Mean Squared Error}(\hat{\theta}) = E((\hat{\theta} - \theta)^2)$$

where $\hat{\theta}$ is the MLE of θ and θ is the true value of θ. We will use the following steps to estimate the bias and mean squared error:

1. Generate 100 random samples from the given distribution with θ = 0.5, 1, 2. We will use the inverse transform method to generate the random samples.

> The inverse transform method is used to generate random samples from a given distribution.
> The algorithm is as follows:
> 1. Generate a random number U from the uniform distribution on the interval $[0,1]$.
> 2. Calculate the inverse of the CDF at the value of U to get the random sample.
> We will use the following CDF to generate the random samples:
> $$F(x) = \begin{cases} \left(1 - e^{-\frac{x}{\theta}}\right)^2 & x > 0 \\ 0 & x \leq 0 \end{cases}$$
> The inverse of the CDF is given by:
> $$F(x) = \left(1 - e^{-\frac{x}{\theta}}\right)^2$$
> $$\sqrt{F(x)} = 1 - e^{-\frac{x}{\theta}}$$
> $$e^{-\frac{x}{\theta}} = 1 - \sqrt{F(x)}$$
> $$-\frac{x}{\theta} = ln(1 - \sqrt{F(x)})$$
> $$x = -\theta ln(1 - \sqrt{F(x)})$$
> We will use the above equation to generate the random samples.

2. Calculate the MLE of θ for each sample by using the method described in Question2(a).
3. Repeat the above steps 1 and 2 for 100 times to get 100 MLEs for each value of θ. Here iteration number is used as seed for random number generation.
4. Calculate the bias and mean squared error of the MLE for each value of θ using the following formulas:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$
$$\text{Mean Squared Error}(\hat{\theta}) = E((\hat{\theta} - \theta)^2)$$

by performing this experiment we the following results $\rightarrow$

1. θ=0.5, mse=0.0011625406930615128, bias=-0.004914924669973003
2. θ=1, mse=0.0046501627722460625, bias=-0.00982984933994567

3. $\theta=2$, mse=0.01860065108898442, bias=-0.019659698679888572

Question no. 3

Given probability density function

$$f(x; \mu, \sigma^2) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(lnx - \mu)^2} & x \geq 0 \\ 0 & otherwise \end{cases}$$

Now using tranformation technique with $Y = Log(X)$
we get:

$f(y) = f(x)|\frac{dx}{dy}|$
where $f(y)$ and $f(x)$ are pdfs.

So,

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(y-\mu)^2}, \quad y \in \mathbb{R}$$

This function looks like a normal distribution with μ and σ as expectation and standard deviation.

Hence we can say that:

If X follows given probability distribution function then Y follows normal distribution function with (μ, σ^2) where $Y = ln(x)$

Hypothesis to test:

$H_0 : (\mu, \sigma) = (\mu_0, \sigma_0)$
$H_1 : (\mu, \sigma) \neq (\mu_0, \sigma_0)$

Our aim is to find test function which we acieve using the likelihood ratio test.

The likelihood function is given by

$$L(\theta) = \prod_{i=1}^{20} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

$$L(\theta) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2}$$

Now Ramdom sample `(Y_1, Y_2,...Y_n)`

Let μ_1 and σ_1 are MLE for μ and σ.

Now the Likelihood ratio is given by;

$$\frac{Sup_{\theta \in \Theta_0} L(\theta)}{Sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)} = \frac{\prod_{i=1}^{n} \frac{1}{\sigma_0\sqrt{2\pi}} e^{\frac{-1}{2\sigma_0^2}(y_i - \mu_0)^2}}{\prod_{i=1}^{n} \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-1}{2\sigma_1^2}(y_i - \mu_1)^2}}$$

$$= (\frac{\sigma_1}{\sigma_0})^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu_0)^2 + \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(y_i - \mu_1)^2}$$

Now as $\sigma_1^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}, \mu_1 = \bar{y}$

$$(\frac{\sigma_1}{\sigma_0})^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu_0)^2 + \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(y_i - \mu_1)^2} < k_1$$

$$\Leftrightarrow (\frac{\sigma_1}{\sigma_0})^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu_0)^2 + \frac{n}{2}} < k_1$$

$$\Leftrightarrow (\sigma_1)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu_0)^2} < k$$

Now more simplification may be not possible theoritically, we use monte carlo to solve further.

The MP level α test is given by

$$\psi(x) = \begin{cases} 1 & (\sigma_1)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu_0)^2} < k \\ 0 & o.w \end{cases}$$

Now k is such that $E_{(\mu,\sigma=\mu_0,\sigma_0)}(\psi(x)) = \alpha$

$$P_{\mu,\sigma=\mu_0,\sigma_0}\left((\sigma_1)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (y_i-\mu_0)^2} < k\right) = \alpha$$

To find the distribution of LHS is quite hard.

Therefore do the following steps to find the value of k.

1. Generate 100000 random samples from Normal distribution with μ_0 and σ_0.
2. Pass this random samples to get the random numbers from LHS's distribution.
3. After getting this 100000 statistic values, we sort it into increasing terms.
4. Now we set k equal to random number that is the index $\lfloor 10000\alpha \rfloor$. which gives us the value for which only $\alpha \cdot 100$ random numbers are less than k $(P(\Lambda(y) < k) = \alpha)$.

To test given data, whether to reject or accept the null hypothesis.

check whether:

$(\sigma_1)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (y_i-\mu_0)^2} < k$

where `y_i` are data points.

In this case we **Accept Our Null Hypothesis**.