MODEL ANSWERS OF MID-SEMESTER LAB EXAMINATION

1. (10 points) Calculate the value (approximate) of

$$\sum_{k=3}^{\infty} \left(\frac{1}{2}\right)^k \ln k$$

using Monte Carlo technique. Mention all the steps including values of different parameters clearly in the report.

Solution:

$$\begin{aligned}\sum_{k=3}^{\infty} \left(\frac{1}{2}\right)^k \ln k &= \left(\frac{1}{2}\right)^2 \sum_{k=3}^{\infty} \left(\frac{1}{2}\right)^{k-2} \ln k \\ &= \frac{1}{4} \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \ln(k+2) \\ &= \frac{1}{4} E\left(\ln(X+2)\right),\end{aligned}$$

where

$$P(X = k) = \left(\frac{1}{2}\right)^k, k = 1, 2, \ldots. \tag{1}$$

ALGORITHM:

1: Generate U from $U(0,\, 1)$ distribution.
2: Set $X = \left\lfloor -\frac{\ln U}{\ln 2} \right\rfloor + 1$ ▷ X has distribution as given in (1)
3: Repeat previous two steps N time to obtain $X_1,\, X_2,\, \ldots,\, X_N$.
4: Return $\frac{1}{N} \sum_{i=1}^{N} \ln\left(X_i + 2\right)$.

The R-software is used to perform the computation taking seed 123. We have taken $N = 10000$. An approximate value of the sum is 2.005.

2. (10 points) Let X and Y be two random variables such that $X \sim N(-1,\, 4)$ and $Y|X = x \sim N(2x + 2,\, 1)$ for all $x \in \mathbb{R}$. Generate 1000 random numbers from the bivariate distribution of $(X,\, Y)$. Calculate the sample means, sample variances and sample correlation coefficient of the bivariate data. Write the steps in the report clearly.

Solution: $X \sim N(-1,\, 4) \implies E(X) = -1, V(X) = 4$. Also, $Y|X = x \sim N(2x + 2,\, 1)$ for all $x \in \mathbb{R}$. Now,

$$E(Y) = E(E(Y|X)) = E(2X + 2) = 2E(X) + 2 = 2(-1) + 2 = 0.$$

and

$$V(Y) = E(V(Y|X)) + V(E(Y|X)) = 1 + V(2X + 2) = 1 + 4V(X) = 17.$$

Also,

$$E(XY) = E(E(XY|X)) = E(X(2X+2)) = 2E(X^2) + 2E(X) = 8.$$

Therefore,

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}} = \frac{8-0}{\sqrt{4 \times 17}} = \frac{4}{\sqrt{17}}.$$

Thus,

$$(X, Y) \sim N_2 \left(\underline{\mu} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & 8 \\ 8 & 17 \end{bmatrix} \right). \tag{2}$$

Using Cholesky factorization, we can write $\Sigma = AA'$, where

$$A = \begin{bmatrix} 2 & 0 \\ 4 & 1 \end{bmatrix}.$$

Therefore, we can use the following algorithm to generate random number from the distribution of (X, Y).

ALGORITHM:

1: Generate U_1 and U_2 from $U(0,\, 1)$ distribution independently.
2: Set $Z_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2)$ and $Z_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$. ▷ $Z_2,\, Z_2 \sim N(0,\, 1)$
3: Set $X = 2Z_1 - 1$ and $Y = 4Z_1 + Z_2$. ▷ $(X,\, Y)$ follows the distribution given in (2)
4: Return $(X,\, Y)$.

All the computations are preformed in R software taking seed as 123. Based on the generated random numbers of size 1000, the sample mean and variance of X are computed as -1.02 and 4.07, respectively. The sample mean and variance of Y are computed as -0.20 and 16.63, respectively. The sample correlation between X and Y is obtained as 0.97.

3. Consider the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \left(1 - e^{-x^2}\right)\left(\frac{\sqrt{2}}{\Gamma(\frac{1}{2})}\int_0^x \frac{e^{-2s}}{\sqrt{s}}ds\right) & \text{if } x \geq 0. \end{cases}$$

(a) (6 points) Generate sample of size 1000 from the above cumulative distribution function. Write all steps clearly in the report.

Solution: Let $X_1 \sim F_1$ and $X_2 \sim F_2$, where F_1, F_2 are CDFs. Also assume that X_1 and X_2 are independent random variables. Then $F(x) = F_1(x)F_2(x)$ is the CDF of $X = \max(X_1, X_2)$. To see it, proceed as follows. For all $x \in \mathbb{R}$,

$$P(X \leq x) = P\left(\max\{X_1,\, X_2\} \leq x\right) = P\left(X_1 \leq x,\, X_2 \leq x\right) = F_1(x)F_2(x).$$

Take

$$F_1(x) = \begin{cases} 1 - e^{-\frac{x^2}{2}} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad F_2(x) = \begin{cases} \frac{\sqrt{2}}{\Gamma(\frac{1}{2})}\int_0^x \frac{e^{-2s}}{\sqrt{s}}ds & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that random numbers from $F_1(\cdot)$ can be drawn using inverse transform technique. Also, $F_2(\cdot)$ is the CDF of $Gamma(0.5, 2)$ distribution, and hence, random numbers from

$F_2(\cdot)$ can be obtained using acceptance-rejection technique. Thus, the following algorithm can be used to generate a random number from the CDF given in the question.

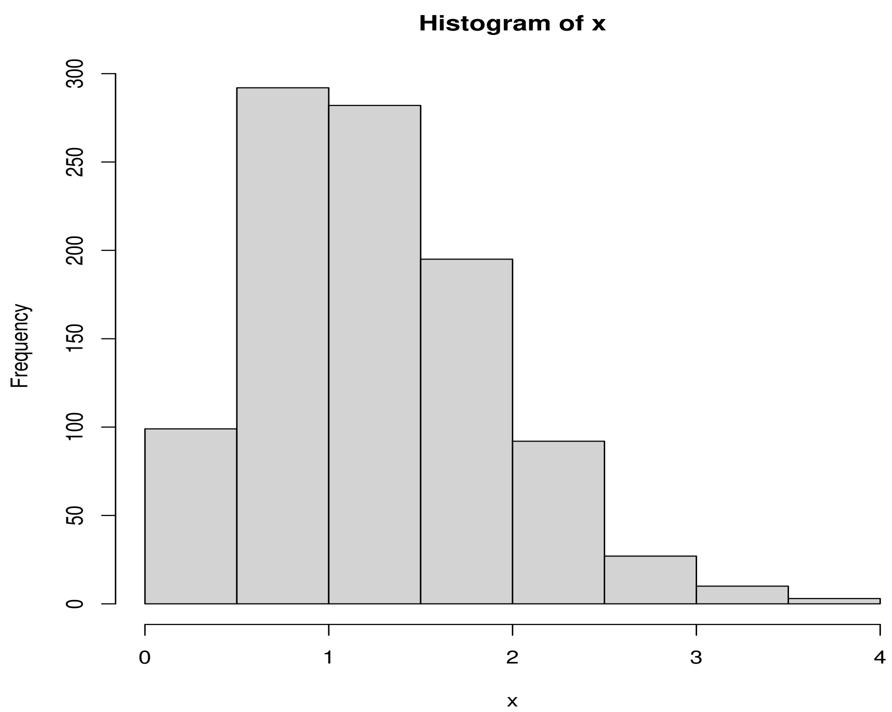
ALGORITHM:

Generate U_1 from $U(0,\, 1)$ distribution.
Set $X_1 = \sqrt{-2 \ln U_1}$. $\triangleright$ X_1 follows $F_1(\cdot)$
repeat
    Generate U_2 from $U(0,\, 1)$.
    if $U_2 < \frac{2e}{1+2e}$ **then**
        Set $X = -\ln\left(2 + \frac{1}{e}\right) - \ln(1 - U_2)$.
    else
        Set $X = \left(1 + \frac{1}{2e}\right)^2 U_2^2$.
    end if
    Generate U_3 from $U(0,\, 1)$.
until
if $X < 1$ **then**
    $e^{-X} \geq U_3$.
else
    $U_3\sqrt{X} \leq 1$.
end if
$X_2 = \frac{X}{2}$. $\triangleright$ X_2 follows $Gamma(0.5,\, 2)$
Return $\max\{X_1,\, X_2\}$.

We use the above algorithm to generate 1000 random numbers from the CDF given in question. For this purpose, we use R software with seed 123.

(b) (2 points) Draw the histogram based on the generated sample in part (a).

Solution: The histogram based on the generated random numbers in the previous part is given bellow. It seems that the distribution is positively skewed.



(c) (2 points) Calculate an approximate value of $E(X)$ based on the sample generated in part (a), where X is a random variable with cumulative distribution function $F(\cdot)$ as given above.

Solution: Once the random numbers are generated from the CDF given in question, $E(X)$ can be approximated by $\frac{1}{N}\sum_{i=1}^{N} X_i$, where X_1, X_2, $\ldots$, X_N are random numbers that are generated. Based on the generated values in part (a), an approximate value of $E(X)$ is 1.26.

[Hint: Try to write F as the CDF of minimum of two appropriate random variables.]