Question 39.2023

Anupama Kulshreshtha
EE22BTECH11009

Let X_1, X_2, ... , X_n be a random sample of size n from a population having probability density function

$$p_X(x;\mu) = \begin{cases} e^{-(x-\mu)}, & \text{if } \mu \leq x < \infty \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\mu \in \mathbb{R}$ is an unknown parameter. If $\hat{M}$ is the maximum likelihood estimator of the median of X_1, then which one of the following statements is true?

A) $\Pr\left(\hat{M} \leq 2\right) = 1 - e^{-n(1-\log_e 2)}$ if $\mu = 1$

B) $\Pr\left(\hat{M} \leq 1\right) = 1 - e^{-n\log_e 2}$ if $\mu = 1$

C) $\Pr\left(\hat{M} \leq 3\right) = 1 - e^{-n(1-\log_e 2)}$ if $\mu = 1$

D) $\Pr\left(\hat{M} \leq 4\right) = 1 - e^{-n(2\log_e 2-1)}$ if $\mu = 1$

Solution: For continuous random variable X, median M is such that,

$$\Pr(X \leq M) = 0.5 \tag{2}$$

The pdf of X is given by,

$$p_X(x) = \begin{cases} e^{-(x-\mu)}, & \text{if } \mu \leq x < \infty \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

Hence, cdf is given by

$$F_X(x;\mu) = \int_{\mu}^{x} e^{-(t-\mu)}dt \tag{4}$$
$$= e^{\mu}[-e^{-x} + e^{-\mu}] \tag{5}$$
$$= 1 - e^{-(x-\mu)} \tag{6}$$

Now,

$$
\begin{aligned}
F_X(x;\mu) &= 0.5 &(7)\\
\implies 1 - e^{-(M-\mu)} &= 0.5 &(8)\\
\implies \hat{M} &= \mu + \ln(2) &(9)
\end{aligned}
$$

Definition 1. *L, the Maximum Likelihood Estimator of the distribution is given by,*

$$
\begin{aligned}
L &= \prod e^{-(x-\mu)} &(10)\\
&= e^{-(\sum x_i - n\mu)} &(11)\\
& &(12)
\end{aligned}
$$

For the Likelihood function to be maximum, $\sum x_i - n\mu$ should be minimum Hence,

$$
\begin{aligned}
X_i &> \mu &(13)\\
\implies \sum x_i &> n\mu &(14)\\
\sum x_i - n\mu &> 0 &(15)\\
\implies \mu &= \frac{\sum x_i}{n} &(16)
\end{aligned}
$$

Given,

$$
\begin{aligned}
p_X(x) &= e^{-(x-\mu)} &(17)\\
& &(18)
\end{aligned}
$$

X_i follows an exponential distribution.

Definition 2. *We know that if,*

$$
\begin{aligned}
p_X(x) &= \lambda_i e^{-\lambda_i x} &(19)\\
S &= X_1 + X_2 + ... + X_n &(20)\\
p_S(n) &= \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} &(21)
\end{aligned}
$$

which is gamma distribution with parameters n and λ

Hence, pdf of

$$
\begin{aligned}
Y &= \sum_{i=1}^{n} x_i &(22)\\
&\text{where,} &(23)\\
p_X(x) &= e^{-(x-\mu)} &(24)
\end{aligned}
$$

will follow gamma distribution with parameter n and $\lambda = 1$, given by,

$$p_Y(x; n, 1) = \frac{x^{n-1}e^{-x}}{(n-1)!} \tag{25}$$

Hence, cdf is given by,

$$F_Y(x; n) = \int_1^x \frac{t^{n-1}e^{-t}}{(n-1)!}\, dt \tag{26}$$
$$= 1 - \Gamma(n, x) \tag{27}$$

where $\Gamma(n, x)$ is incomplete gamma function. Thus,

$$\Pr\left(\hat{M} \le k\right) = \Pr\left(\frac{\sum x_i}{n} + \ln(2) \le k\right) \tag{28}$$
$$= \Pr\left(Y/n + \ln(2) \le k\right) \tag{29}$$
$$= \Pr\left(Y \le n(k - \ln(2))\right) \tag{30}$$
$$= F_Y(n(k - \ln(2))) \tag{31}$$
$$= 1 - \Gamma(n, n(k - ln2)) \tag{32}$$

It needs a value of n to be computed.

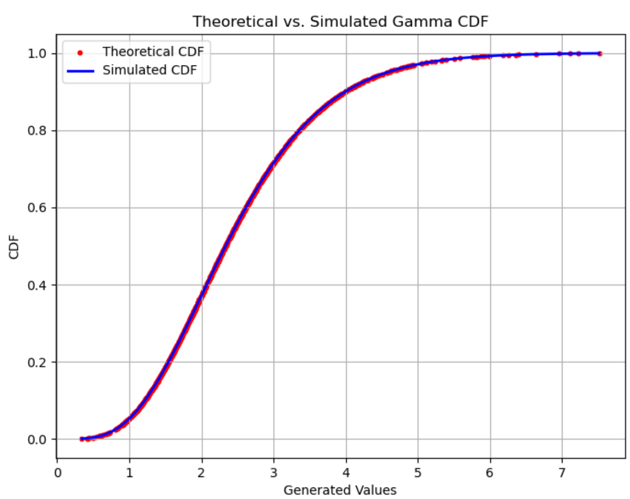

Figure 1: Verifying gamma cdf through simulation

Steps for simulation in C:

1. Import the necessary libraries, including 'stdio.h','stdlib.h' and 'math.h'.

2. Write functions for generating exponential distribution, gamma pdf and gamma cdf.

3. In the main function, the exponentials generated using the function are added and stored in variable 'gamma samples' for each sample, to get sum of exponentials.

4. Then the simulated gamma cdf values are calculated using 'gamma cdf' function and the calculated sum of exponentials.

5. The code is then compiled using GCC compiler in the terminal (gcc simulation.c -o simulation -lm), and the results are stored in an output.txt file. (./simulation>output.txt)

6. The output file is loaded into the python code with theoretical cdf values, and the final graph between theoretical and simulated cdf is plotted.

7. The simulated and theoretical cdf values match, which verifies the gamma cdf through simulation.