Monte Carlo Simulation - MA 323

Lab 09 - Report - Lakshya Kohli - 210123077

Answer a.

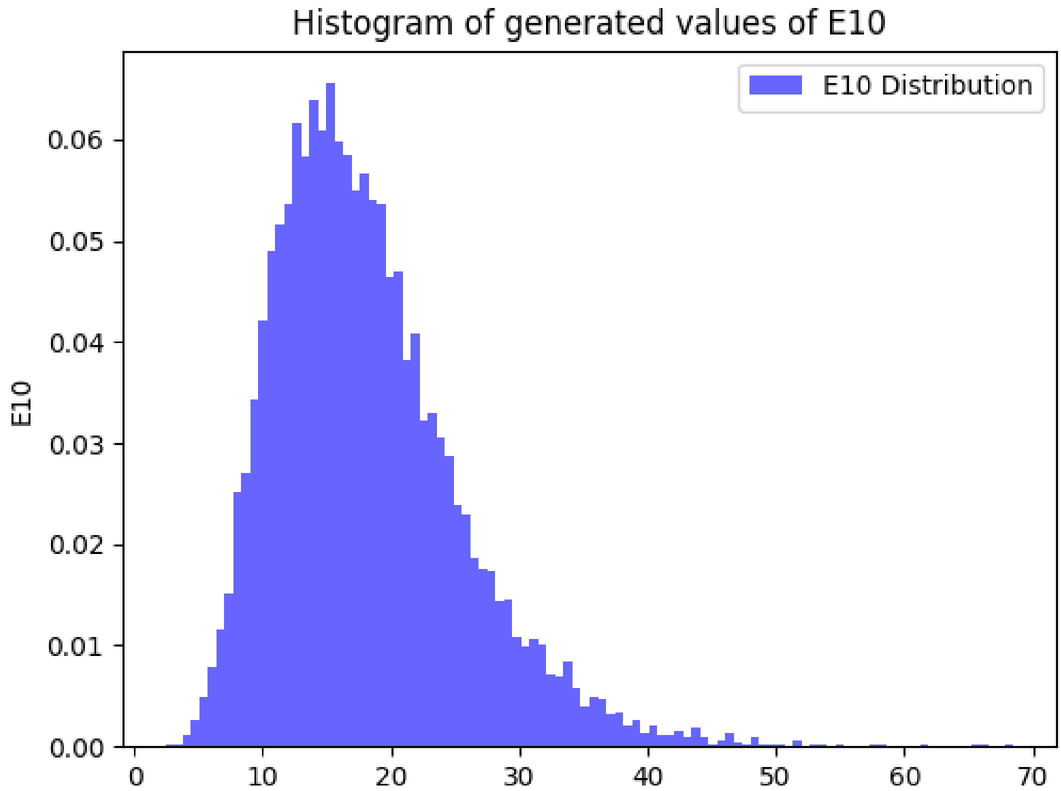
E_{10} in terms of T_j, j = 1, 2,..., 10:

$$E_1 = T_1$$
$$E_2 = T_1 + T_2$$
$$E_3 = T_1 + T_3$$
$$E_4 = T_1 + T_2 + T_4$$
$$E_5 = T_1 + T_2 + T_5$$
$$E_6 = T_1 + T_3 + T_6$$
$$E_7 = T_1 + T_3 + T_7$$
$$E_8 = T_1 + T_3 + T_8$$
$$E_9 = \max\{T_1 + T_2 + T_5,\ T_1 + T_3 + T_6,\ T_1 + T_3 + T_7\} + T_9$$
$$E_{10} = \max\{E_4,\ E_8,\ E_9\} + T_{10}$$
$$\Downarrow$$
$\star$
$$E_{10} = \max\{T_1 + T_2 + T_4,\ T_1 + T_3 + T_8,\ \max\{T_1 + T_2 + T_5,\ T_1 + T_3 + T_6,\ T_1 + T_3 + T_7\} + T_9\} + T_{10}$$

Answer b.

Approximate value of mean of E_{10} using a simple Monte Carlo is 18.157128017032395.

Answer c.



The above histogram is expected as E10 is the sum of independently distributed exponential functions with different mean, which will result in gamma distribution. The above histogram is ***positively*** skewed.

Mathematically, if we have n independent exponential random variables X_1 , X_2 , ..., Xn with means ($1/\lambda_1$), ($1/\lambda_2$), ..., ($1/\lambda n$), respectively, then their sum $Y = X_1 + X_2 + \ldots + X_n$ will be gamma-distributed as:

$Y \sim \text{Gamma}(\alpha, \beta)$, where:

α (shape parameter) = n

β (scale parameter) = $1 / (\lambda_1 + \lambda_2 + \ldots + \lambda n)$

Answer d.

Approximate value of the probability that the project misses the deadline using a simple Monte Carlo is 0.0.

This is even clearly evident from the histogram generated in part c, that the samples of E10 have very less probability in regions with value more than 70, so is the observed probability zero.

Even if I increase my sample size to say 1 million, then the probability of missing the deadline is negligible and is in the order of 10^{-5}.

The observed standard deviation of E10 = 7.329207216346555

The performance in context of the deadline is good as observed by the missing probability. This is due to the skewness of the above histogram which depicts that the value of E10 in the region with the value greater than 70 is very low.

Now, in the subsequents parts we are using importance sampling so as to get samples from the above region(>=70), with very less probability, or zero.

Answer e.

Approximate value of the probability that the project misses the deadline using the importance sampling technique is 0.472. The reason for increment of this probability is the underlying importance sampling technique used.

The observed standard deviation of E10 = 28.883882856980716

Effective Sample Size = 643.8769098997877

Answer f.

K Value	Standard Deviation	Effective Sample Size	Probability Missing deadline
3.0	28.54346329821982	659.3267066030155	0.4781
4.0	28.446791636963177	663.81554391516	0.4726
5.0	28.767133205611298	649.1137852965588	0.4716

Answer g.

Standard Deviation	Effective Sample Size
28.883882856980716	643.8769098997877

K Value	Standard Deviation	Effective Sample Size
3.0	28.54346329821982	659.3267066030155
4.0	28.446791636963177	663.81554391516
5.0	28.767133205611298	649.1137852965588

Answer h.

```
k = 5 has the minimum effective sample size.
99% Confidence interval (71.88893992872913, 73.37332400213866)
```

I have used a seed, though by removing that seed I have got k=3, k=4, k = 5, all three values which have the minimum effective sample size.