Q.1 (a) The three case study question (i) Check-out counter at airport (ii) Cash counter analysis in Bank and (iii) Customer flow analysis in super market.

Solution: Given below are two case studies. The CUSTOMER FLOW ANALYSIS IN SUPER MARKET case study can be taken as it is as guide for the subquestion (iii). The second case study below, CHECK-OUT COUNTER SIMULATION AT SUPER MARKET can be taken as a guide for both sub question (i) as well as (ii) above with very trivial changes.

*[**Guidelines to the Evaluator**: 1) as case study answer would obviously be very different from one student to next the examiner i points to consider are: i) whether the student is able to properly describe the problem statement is supposed to use his/her discretion to give fair marks to the student. The important and explicitly mention the objectives and purpose of the simulation study. ii) Whether the student is able to identify important events for his chosen case study problem iii) what kind of random input elements the student is able to identify for the case study. iv) What probability distribution does he select for the random inputs in his simulation and does he justify his choice of the distribution convincingly. V) What measures of performance does he has used to measure the performance of his system under simulation? Does the measures of performance proper for the purpose objectives of the case study and finally vii) what conclusion is drawn by the student post the simulation run.*

2) The marks distribution can be as follows: 2 marks for problem statement and stating the objectives of the case study. 3 marks for the model description. 2 marks for pre simulation analysis, 2 marks for post simulation findings and 1 mark for appropriate conclusion]

Example sample answer can be as follows: [Students may or may not take the sample probability distributions. Marks to be given as per the explanation provided.]

(iii)CUSTOMER FLOW ANALYSIS IN SUPER MARKET

Problem Definition:

A major supermarket wants to open its new branch in the neighbouring city. It needs to finalize the location and other details of the same. Therefore to assist its decision, simulation is required to analyse the current flow of customer in the current super market. So the output of the simulated model can be used as inputs in the decision such as which market area should be targeted, should there be more than 3 billing counters and so on. Simulation would be best suited for this situation since it would throw light on the crucial parameters such as the amount of time the customer waits in the queue, the amount of time the billing counter is idle and so on.

Description of the simulated problem:

The single server model of simulation best resembles this situation. The simulated model would require input parameters such as the inter-arrival time of the customers, and the amount of time it takes to serve each customer. Here the customers would be served on first come first serve basis. If the arrived customer is the first one then he is served immediately if not then he must await his turn.

The sample probability distribution of inter-arrival times:

Inter-arrival Time (minutes)	**Probability**	**Cumulative Probability**	**Random Digit Assignment**
1	0.125	0.125	000-125
2	0.125	0.250	126-250
3	0.125	0.375	251-375
4	0.125	0.500	376-500
5	0.125	0.625	501-625
6	0.125	0.750	626-750
7	0.125	0.875	751-875
8	0.125	1.000	876-000

The sample service Time probability distribution:

Service Time (minutes)	**Probability**	**Cumulative Probability**	**Random Digit Assignment**
1	0.10	0.10	01-10
2	0.20	0.30	11-30
3	0.30	0.60	31-60
4	0.25	0.85	61-85
5	0.10	0.95	86-95
6	0.05	1.00	96-00

Sample simulation table can be constructed as below:

Customer	**Inter-arrival Time**	**Arrival Time**	**Service Begin Time**	**Service Time**	**Service End Time**	**Waiting Time in Queue**	**Time customer spends in market**	**Idle time of the counter**
1	-	0	0	4	4	0	4	
2	1	1	4	2	6	3	5	0
3	1	2	6	5	11	4	9	0
4	6	8	11	4	15	3	7	0

5	3	11	15	1	16	4	5	0
6	7	18	18	5	23	0	5	2
7	5	23	23	4	27	0	4	0
8	2	25	27	1	28	2	3	0
9	4	29	29	4	33	0	4	1
10	1	30	33	3	36	3	6	0

Average Waiting Time = $\frac{\text{total time customers wait in queue}}{\text{total no. of customers}}$ = 19/10 = 1.9(minutes)

This means that on an average the customer had to wait for 1.9 minutes.

So this waiting time must be reduced to manage the flow of the customers.

Pre-simulation Analysis:

In order to start simulating the problem the statistics must be collected about how frequently the customer's arrive and approximately how much time does it take to serve one customer. Once these statistics are collected the various frequent scenarios can be modelled and how the system would react to them can be determined.

Post-simulation Analysis:

Once the model is built and several iterations of simulations are done, then the output can be used to determine the flow of the customers in the super market. The parameters calculated from simulation would be of paramount importance in determining the number of counters required in the new market, the appropriate size of the market and so on. The total time spent by the customers in the market would give a rough estimate of how much time one customer would approximately stay in the market and when would the next customer arrive.

Conclusion:

Hence simulating the customer flow would be the best solution to draw inferences on the pattern of the arrival of the customers. And additionally the model would also safe guard the new market against some of the unforeseen problems.

CHECK-OUT COUNTER SIMULATION AT SUPER MARKET (Airport checkout counter can be similar)

Problem Definition:

It has been lately observed the rush at the supermarket is consistently increasing. It is usually at its peak in the evenings. A restructuring of the checkout counters is now required to handle the mass efficiently. In order to carry out the restructuring of the counters an estimate is required whether as to add additional counters or improve the processing speeds at the existing counters. In order to come to a decision it is necessary to simulate the current scenario, so that the output of the simulated model can be analysed in order to come to an appropriate restructuring decision.

Description of the simulated model:

The supermarket checkout counter tends to follow the dual-server queuing system in which there are two counters available to serve the customers. Hence the model follows a multiple server model with an infinite queue. So the simulated model would encompass counters that are waiting for an incoming customer. When the customer arrives at the checkout counter he is attended on the first come first serve basis. If the arrived customer is the first customer in the queue the he is served immediately of if there are other customers in the queue then the customer must await his turn. Initially the server/ checkout counter is in idle state, when the first customer arrives the server goes in processing state. The inter-arrival time between two customers results in increasing the idle time for the server.

The probability distribution is given for the inter-arrival time of the customer:

Time Between Arrival	Probability	Cumulative Probability	Random Digit Assignment
1	0.05	0.05	01-05
2	0.25	0.30	06-30
3	0.30	0.60	31-60
4	0.25	0.85	61-85
5	0.15	1.00	86-00

The probability for Counter1's service time:

Counter1's service time	Probability	Cumulative Probability	Random Digit Assignment
2	0.20	0.20	01-20
3	0.35	0.55	21-55
4	0.30	0.85	56-85
5	0.15	1.00	86-00

The probability for Counter2's service time:

Counter1's service time	Probability	Cumulative Probability	Random Digit Assignment
3	0.38	0.38	01-38
4	0.25	0.63	39-63
5	0.18	0.81	64-81
6	0.19	1.00	82-00

A sample simulation table can be constructed for the model as below, it is assumed that when counter1 and counter2 are both idle the customer will choose counter1 and only if the counter1 is busy when a new customer arrives only then the processing is handed over to counter2:

Customer Number	RD for inter-arrival time	Inter-arrival Time	(Clock) Arrival	RD for service time	Counter 1				Counter 2			
					Service Begin Time	Service Time	Service End Time	Idle Time	Service Begin Time	Service Time	Service End Time	Idle Time
1	-	0	0	25	0	3	3	-	-	-	-	-
2	95	5	5	38	5	3	8	2	-	-	-	-
3	26	2	7	61	-	-	-	-	7	4	11	-
4	43	3	10	52	10	3	13	2	-	-	-	-
5	71	4	14	01	14	2	16	1	-	-	-	-
6	12	2	16	47	16	3	19	0	-	-	-	-
7	10	2	18	22	-	-	-	-	18	3	21	7
8	55	3	21	12	21	2	23	2	-	-	-	-
9	62	4	25	10	25	2	27	2	-	-	-	-
10	44	3	28	54	28	3	31	1	-	-	-	-

Average Service Time and waiting time can be calculated as:

$$\text{Average Service Time (minutes)} = \frac{\text{Total Service Time (minutes)}}{\text{Total Number of Customers}} = 28 / 10 = 2.8 \text{ minutes}$$

$$\text{Average waiting time of customers} = \frac{\text{Total Time spent in queue}}{\text{Total no. of customers who wait}} = 0/0 = 0 \text{ minutes}$$

It means that in a sample of 10 customers none of the customers had to await their turn of processing.

Pre-simulation Analysis

In order to start with the simulating the problem certain analysis must be done in order to collect some statistics about the problem so that the model can be used to simulate the most frequent scenarios that the original system might have to handle. The most important statistic that would be required is the observed inter-arrival frequency of the customers. The approximate inter-arrival time between the incoming customers should be noted since this would be the major input for the model.

The second important statistic is the service time, i.e. the amount of time the person at the check-out counter takes to serve a single customer. This is necessary since we need to determine the amount of time a customer spends waiting in the line. And finally we need to compute the amount of time for which the counter was idle, in order to improve the overall productivity.

Post-Simulation Analysis

Once the process of simulating the various possible scenarios is completed, we can observe the various parameters calculated. These parameters would play a crucial role in making the counter restructuring decision. The total amount of time spent by the customers waiting in the line and the total idle time of the server would provide inputs for determining the efficiency of the person serving the customer at the counter. The inter-arrival time of the customers would suggest the pattern or frequency of the customer arrival. This would help to manage the huge population at the peak hours.

Conclusion

Thus it can be concluded that simulation is of vital importance in making the real-life decisions that would have an impact on the current system. Simulating a scenario is just as good as preparing for the unforeseen problems; it gives the power and ability to tackle the problems on the real system efficiently.

Q.1 (b) Suggest a distribution for the following process in the computer assembly shop

- **(i) Number of defective chips found in a lot of n chips**
- **(ii) Number of computer chips that we must inspect to find 5 defective chips**
- **(iii) Time to assemble a computer which is the sum of the times required for each assembly operation**
- **(iv) Time to failure for a disk drive.**
- **(v) If the minimum, most-likely and maximum time required to test a product is known.**

Solution: (i): Assuming that the lot of n chips is randomly selected, one by one, from a bigger lot of size N, we can use binomial distribution for the random variable X, representing the no of defective chips found in a sample of n chips.

However, if the sample size n is relatively large as compared to bigger lot size N, we should not use binomial distribution because in such a case the individual trials, where a single chip is randomly selected, are not independent of each other, violating one of the basic assumptions necessary for trials to be Bernoulli trials.

Further, assuming that binomial distribution is a reasonable choice, if the sample size n is very large (say, $n > 100$) and probability p of finding a defective chip in an individual trial is very small (say, $p <= 0.05$) then a Poisson distribution can be used as an approximation for binomial distribution with $\propto = n \times p$

(ii): Since negative binomial distribution is the distribution of the number of trials until the k^{th} success, for k = 1, 2,… negative binomial distribution should be used with parameters k=5 and ‘p’ as probability of finding a defective chip on any single trial.

(iii): Normal distribution should be used to represent time to assemble a computer. This is because Normal distributions are usually used to model a process that is derived from the sum of a large number of independent random variables. And our process of assembling a computer which consists of many individual assembly operations is exactly such a process.

(iv): Since Weibull distribution is used to model the time until failure of a piece of component, **Weibull distribution should be used to represent time to failure of a disk drive.**

(v): Triangular distribution is usually used to model a process when only the optimistic, most likely and pessimistic values of the distribution are available. Since in our case, the minimum, most-likely, and the maximum time required to test a product is known **we should use <u>triangular distribution</u> for the random variable of the time required to test a product.**

*[**Guidelines to the Evaluator**: **i)** allot 1 mark for each sub question.*

__ii)__ If the student has given accurate answers for each sub question along with appropriate justifications give full 5 marks.

__iii)__ If only accurate answers are given without justifications give 3 marks.

__iv)__ For the in-between cases use your fair sense of judgement to give appropriate marks]

Q. 1 (c) Use the Multiplicative Congruential method to generate sequence of five three digit random numbers using $x_0 = 37$, a = 7, m = 1000

Solution: In step 1 we generate sequence of five random integers X_i's i = 1 to 5 using the general formula of multiplicative Congruential generator as follows:

$X_i = aX_{i-1} \bmod m$; i = 1 to 5 eq (1)

Since the multiplier *a* and the modulus *m* are given as 7 and 1000, respectively we have

$X_i = 7X_{i-1} \bmod 1000$; i = 1 to 5 eq 2

We start the generation process by feeding the above generator with the given initial seed value $X_0 = 37$ to get X_1 and then repeating the process recursively. Thus we get

$X_1 = aX_0 \bmod m;$
$= (7*37) \bmod 1000$
$= 259 \bmod 1000$
$= 259$

$X_2 = 7 * X_1 \bmod 1000$
$= 7 * 259 \bmod 1000$

$= 1813 \bmod 1000$
$= 813$

$X_3 = 7 * X_2 \bmod 1000$
$= 7 * 813 \bmod 1000$
$= 5691 \bmod 1000$
$= 691$

$X_4 = 7 * X_3 \bmod 1000$
$= 7 * 691 \bmod 1000$
$= 4837 \bmod 1000$
$= 837$
$X_5 = 7 * X_4 \bmod 1000$
$= 7 * 837 \bmod 1000$
$= 5859 \bmod 1000$
$= 859$

In step 2 we convert the above 5 random integers X_i's into the required three digit random numbers R_i's by computing with the following formula:

$R_i = X_i/m \quad i = 1 \text{ to } 5$

Thus we get:
$R_1 = X_1/1000 = 259/1000 = 0.259$

$R_2 = X_2/1000 = 813/1000 = 0.813$

$R_3 = X_3/1000 = 691/1000 = 0.691$

$R_4 = X_4/1000 = 837/1000 = 0.837$

$R_5 = X_5/1000 = 813/1000 = 0.859$

Thus the sequence of five three digit random numbers is:
0.259, 0.813, 0.691, 0.837, 0.859

*[**Guidelines to the Evaluator**: **i)** Give full marks if the method is correct and the answers are also accurate **ii)** In case the answers are incorrect due to silly mistakes but if the method is correct, you might give 4 marks **iii)**If the method is even partially correct give 2 to 3 marks based on your judgement.]*

2 (a) By using Inverse Transform Technique which of the distribution's random variates can be generated?

Develop a random-variate generator for a random variable X with pdf:

$$\mathbf{f(x) = e^{2x}}, \quad -\infty < \boldsymbol{x} \le 0$$
$$\mathbf{= e^{-2x}}, \quad 0 < \boldsymbol{x} < +\infty$$

Generate random-variate for 0.23, 0.81 and 0.37

Solution:

The inverse-transform technique, in principle, can be used to generate random variates for any distribution. But in practice it is more useful if the cdf F(x) of the distribution is of a simple form such that it is easy to derive its inverse cdf F^{-1} from it. (In technical jargon we say that the inverse-transform technique is more useful if the continuous distribution has closed form inverse cdf..)

Thus generally the inverse-transform technique is used to generate random variates of the following continuous distributions:

1) exponential,
2) uniform,
3) Weibull,
4) triangular and
5) Empirical continuous distributions,

Inverse-transform technique is also used to generate random variates of many discrete distributions like:

1) discrete uniform,
2) geometric
3) Discrete empirical distributions using the table lookup procedure.

Developing random-variate generator for random variable X with the given pdf

The following step by step procedure can be used to develop a random-variate generator for a random variable X with the given pdf

<u>Step 1: Compute the cdf of the random variable X from its pdf:</u>

For our distribution since the pdf is given as:

$$f(x) = e^{2x}, \quad -\infty < x \le 0$$
$$= e^{-2x}, \quad 0 < x < +\infty$$

We have the cdf defined differently for different ranges of X as follows:

For $-\infty < X \le 0$	For $0 < X < \infty$

$F(X) = \int_{-\infty}^{x} e^{2t}dt$ $= \left[\frac{e^{2t}}{2}\right]_{-\infty}^{x}$ $= \frac{e^{2x}}{2}$	$F(X) = \int_{-\infty}^{0} e^{2t}dt + \int_{0}^{x} e^{-2t}dt$ $= \left[\frac{e^{2t}}{2}\right]_{-\infty}^{0} + \left[-\frac{e^{-2t}}{2}\right]_{0}^{x}$ $= \frac{1}{2} + \frac{1}{2} - \frac{e^{-2x}}{2}$ $= 1 - \frac{e^{-2x}}{2}$

Thus the cdf F(X) for the given distribution is:

$$F(X) = \begin{cases} e^{2x}/2\ , & -\infty < X \leq 0 \\ 1 - e^{-2x}/2\ , & 0 < X < +\infty \end{cases}$$

Step 2: Set $F(X) = R$ on the range of X, that is on, $-\infty < X < +\infty$

For our distribution it becomes:

$$\frac{e^{2x}}{2} = R, \qquad -\infty < X \leq 0$$

$$1 - \frac{e^{-2x}}{2} = R, \qquad 0 < X < +\infty$$

Step 3: Solve $F(X) = R$ for X in terms of R to obtain $F^{-1}(R) = X$

In our case we do it as follows:

For $0 < R \leq 1/2$ $\frac{e^{2x}}{2} = R$ $e^{2x} = 2R$ $2x = \ln(2R)$ $x = \frac{1}{2}\ln(2R)$	For $\frac{1}{2} < R < 1$ $1 - \frac{e^{-2x}}{2} = R$ $\frac{e^{-2x}}{2} = 1 - R$ $e^{-2x} = 2 - 2R$ $-2x = \ln(2 - 2R)$ $x = -\frac{1}{2}\ln(2 - 2R)$

Thus we have a random-variate generator that can generate random variate X_i's given random numbers R_i's as follows:

$$X_i = \begin{cases} 1/2\ \ln(2R_i)\ , & 0 < R_i \leq 1/2 \\ -1/2\ \ln(2 - 2R_i), & \frac{1}{2} < R_i < 1 \end{cases} \qquad (1)$$

Generating random-variates for sequence of random numbers 0.23, 0.81, 0.37

Let $R_1 = 0.23, R_2 = 0.81, and\ R_3 = 0.37$

We want to generate corresponding random variates X_1, X_2, X_3 by applying random-variate generator of equation (1) above. This can be done as follows:

Since $R_1 = 0.23 < 1/2$ by applying equation (1) appropriately, we get:

$$X_1 = \frac{1}{2}\ln(2\times 0.23),$$

$$X_1 = \frac{1}{2}\ln(0.46) = -0.388$$

Since $R_2 = 0.81 > 1/2$ by applying equation (1) appropriately, we get:

$$X_2 = -\frac{1}{2}\ln(2 - 2 \times 0.81) = -\frac{1}{2}\ln(2 - 1.62) = -\frac{1}{2}ln(0.38) = 0.484$$

And lastly, since $R_3 = 0.37 < 1/2$ by applying equation (1) appropriately, we get:

$$X_3 = \frac{1}{2}\ln(2 \times 0.37) = \frac{1}{2}\ln(0.74) = -0.150$$

Thus random-variates for 0.23, 0.81 and 0.37 are -0.388, 0.484 and -0.15, respectively

*[**Guidelines to the Evaluator**: The solution to this question has 3 parts; part 1 is a theory part which asks random variates of which distributions can be generated using inverse transform method, part 2 asks to develop random variate generator for the problem distribution whose pdf is given and part 3 asks to actually generate random variates from the given random numbers by applying the developed random variate generator in part 2*

Marks allotment:

- ***i)*** *Internally you can allot 3 marks to the theory part, i.e part 1, and remaining 5 marks for the parts 2 and 3, the problem solving parts.*
- ***ii)*** *Within the 5 marks allotted to parts 2 and 3, 4 marks can be allotted to part 2 and only 1 mark can be allotted to part 3*
- ***iii)*** *In part 1 you can give full 3 marks if the distributions are listed along with proper justification. You can cut 1 mark if justifications are not given.*
- ***iv)*** *In part 2 cut just 1 mark for silly mistakes. In such a case give 3 marks instead of allotted 4 to part 2*
- ***v)*** *Use your sense of fair judgement for in-between cases.*

]

Q.2 (b) What do you understand by model verification and validation? How would you validate input-output transformation of a model?

Solution: *[**Guidelines For Solution: i)** For the model verification and validation part refer Jerry Banks 5th editions chapter 1, section 12: Steps in simulation study as well as you can refer initial pages of chp 10: Verification,Calibration and Validations of Simulation Models*

ii) *For the input-output transformation part refer chapter 3, section 3.3: Validating Input-Output Transformations*

*[**Guidelines regarding marks allotment**: First part of the question, about verification and validation, can be allotted 3 marks and the second part of the question, regarding input-output transformation, can be allotted 4 marks]*

Q.3 (a) What is modelling and simulation? Explain different types of simulation models with example.

Solution: *[**Guidelines for solution:** This is a straightforward descriptive question the solution for which can be found in chapter 1 of Jerry Banks book, 5th edition, specifically the answer for the first part is on the 1st page in the introductory paragraph. For the second part of the question with respect to type of models the answer is again in chapter 1 section 10: Types of Models]*

*[**Guidelines for allotment of marks:** Part 1 of the question can be allotted 3 marks and part 2 can be allotted 5 marks of the total 8 marks for the question]*

Q.3 (b) Suppose that arrivals to post office occurs at a rate of 2 per minute from 8 A.M until 12 P.M, then drops to 1 every 2 minutes until day ends at 4 P.M. What is the number of arrivals between 11 A.M and 8 P.M?

Solution:

The rate of arrivals is 2 per minute from 8 A.M to 12 P.M & ½ per minute from 12 to 4 P.M.

Since the post office closes at 4 P.M the arrival rate for the period 4 P.M to 8 P.M can be considered as 0 per minute.

Since we are measuring the time t in hours, we should convert the rate to per hour as under:

$$\lambda(t) = 60\, mins\ \times 2\ per\ min = 120 \qquad 0 \le t < 4$$

$$= 60\, mins \times \frac{1}{2}\ per\ min = 30 \qquad 4 \le t \le 8$$

$$= 0 \qquad 8 < t \le 12$$

Where $t = 0$ corresponds to 8 A.M, $t = 1$ corresponds to 9 A.M, and so on.

The arrival process to post office can be said to be a Non Stationary Poisson Process (NSPP) since the rate of arrival varies in different time periods.

Thus the expected number of arrivals by time t is as given below:

$$\Lambda(t) = \int_0^t \lambda(s)\, ds\ = \int_0^t 120s\, ds\ \ = \mathbf{120}\boldsymbol{t} \qquad \mathbf{0 \le}\ \boldsymbol{t} \mathbf{< 4}$$

$$= \int_0^4 120s\, ds + \int_4^t 30s\, ds$$

$$= \mathbf{30}\boldsymbol{t} + \mathbf{360} \qquad \mathbf{4 \le} \boldsymbol{t} \mathbf{\le 8}$$

$$= \mathbf{0} \qquad \mathbf{8 <} \boldsymbol{t} \mathbf{\le 12}$$

We want to compute the expected number of arrivals between 11 A.M and 8 P.M.

But since the post office closes by 4 P.M, the duration from 4 P.M to 8 P.M need not be considered as it would not contribute any arrivals to the total expected arrivals during the period.

Thus in effect we require computing the expected number of arrivals from 11 A.M to 4 P.M only.

Since 11 A.M corresponds to $t_1 = 3$ and 4 P.M corresponds to $t_2 = 8$ the expected number of arrivals can be given as:

$$\begin{aligned}\Lambda(t_2) - \Lambda(t_1) &= (30t_2 + 360) - (120t_1)\\ &= (30 \times 8 + 360) - (120 \times 3)\\ &= 600 - 360\\ &= 240\end{aligned}$$

Conclusion: The post office can expect 240 arrivals between 11A.M and 8 P.M

*[**Guidelines to the Evaluator: i)** Due to silly mistake if the answer is incorrect but the basic method is sound you might cut maximum 1 to 2 marks based on your discretion. **ii)** Note that the student might have used different method from what is used here. If he is getting the same correct answer by some other method, if you find that the method is perfectly sound and logical then you can give him full marks]*

Q.4 (a) auto correlation question

Solution 4 (a):

For the subsequence to be tested the values for the parameters subsequence start location i, lag l, total length of the complete sequence N, and the largest integer M, such that $i + (M + 1) \times l \leq N$ is as given below:

$i = 3$ (Since the subsequence to be tested for auto-correlation begins with the 3rd number)

$l = 5$ (Since the subsequence to be tested for auto-correlation includes every 5th number)

$N = 30$ (Since there are in all 30 numbers in the complete sequence given)

$M = 4$ (Since the largest integer M, such that $i + (M + 1) \times l \leq 30$ is 4)

Now let's start to test the subsequence for auto-correlation as follows:-

Step 1: Establishing null hypothesis H_0 and alternative hypothesis H_1

H_0: The auto-correlation ρ_{il} is zero for the subsequence.
Mathematically we say that $\rho_{35} = 0$

H_1: The auto-correlation is non-zero for the subsequence. Mathematically we say this as, $\rho_{il} \neq 0$. That is, in our case, $\rho_{35} \neq 0$

Step 2: Computing the test statistic Z_0

$Z_0 = \frac{\widehat{\rho_{il}}}{\sigma_{\widehat{\rho_{il}}}}$ is used as the test statistic for the auto-correlation test where $\widehat{\rho_{il}}$, is given as:

$\widehat{\rho_{il}} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+kl} R_{i+(k+1)l}\right] - 0.25$

And the standard deviation $\sigma_{\widehat{\rho_{il}}}$ is given as:

$\sigma_{\widehat{\rho_{il}}} = \frac{\sqrt{13M+7}}{12(M+1)}$

Since for our subsequence $i = 3, l = 5, \& \ M = 4$ the above becomes:

$\widehat{\rho_{35}} = \frac{1}{4+1}\left[\sum_{k=0}^{4} R_{3+5k} R_{3+5(k+1)}\right] - 0.25$

$= \frac{1}{4+1}[0.23 \times 0.28 + 0.28 \times 0.33 + 0.33 \times 0.27 + 0.27 \times 0.05 + 0.05 \times 0.36] - 0.25$

$= -0.1945$

and

$\sigma_{\widehat{\rho_{35}}} = \frac{\sqrt{13(4)+7}}{12(4+1)}$

Therefore the test statistic assumes the value

$Z_0 = -\frac{0.1945}{0.1280} = -1.516$ (1)

Step 3: Establishing the test criteria using the appropriate significance level α

In auto-correlation test we do not reject the null hypothesis H_0 if the test statistic Z_0 satisfies the following expression:

$-Z_{\propto/2} \leq Z_0 \leq Z_{\propto/2}$

Where $\propto$ is the level of significance and $Z_{\propto/2}$ is the critical value for the given level of significance $\propto$

In our problem it is given that level of significance is 0.05 and critical value is 1.96.

Therefore in our case:

$\propto = 0.05$ and $Z_{\propto/2} = Z_{0.025} = 1.96$

Therefore we would not reject the null hypotheses if

$-1.96 \leq Z_0 \leq 1.96$ (2)

Step 4: Applying the test criteria to the test statistic and concluding

Since in step 2 we have computed $Z_0 = -1.516$ and since $-1.96 \leq Z_0 = -1.516 \leq 1.96$

Therefore in this case, Z_0 satisfies the inequality given in (2) and thus we do not reject the null hypotheses H_0 that claims that the autocorrelation for the subsequence is zero.

Conclusion: The hypothesis of independence (i.e zero autocorrelation) cannot rejected on the basis of this test.

[Note: This is a solved example in Jerry Banks and John Carlson's book: Discrete Event System Simulation, 5th edition, See Example 8, Chapter 7: Random Number Generation under Section 4.2: Tests for Autocorrelation]

*[**Guidelines to the Evaluator**: **i)** Even if the student has not formulated the solution in this ideal step by step manner of establishing the two hypothesis and then proceeding in the systematic manner full marks can be given if the basic method logic is correct and conclusion is also not wrong. **ii)** If the method is correct but the conclusion is incorrect you can cut 1 to 2 marks based on your discretion of whether the conclusion went wrong because of some silly mistake or because of deficient understanding on part of the student.]*

Q.4 (b) Which tests are used to test "Goodness of fit". Describe anyone of them

*[**Guidelines to the Evaluator**: **i)**The student is expected to list all the tests for testing "Goodness of fit" with a little brief description of what is the purpose of such tests in general. For the second part the student is expected to explain one of the listed test in a little more detail **ii)** 3 marks can be allotted to part 1 and 4 marks can be allotted to part 2 out of the total 7 marks for the question]*

Q.5 (a) Prove that "if arrival process follows Poisson distribution then inter arrival time follows exponential distribution".

If the mails coming at a server are in accordance follows Poisson Process, with the mean rate of one every 36 hours. Find the probability that the next mail will come between 24 and 48 hours after the last mail.

Solution:

<u>To prove that: "if arrival process follows Poisson distribution then inter arrival time follows exponential distribution"</u>

It is given that the arrival process is a Poisson Process. Therefore, let's assume the following parameters for the Poisson Process:

λ: the mean rate of arrival of the Poisson Process.

$N(t)$: the counting function for the Poisson Process that counts the number of arrivals in time duration [0,t], for all $t \geq 0$.

Since the arrival process is given to be a Poisson Process with mean rate λ and counting function $N(t)$, the probability that N(t) is equal to n (i.e the number of arrivals is equal to n), is given by the pdf of the Poisson distribution as shown below:

$$P[N(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \qquad \text{For all } t \geq 0 \ \& \ n = 0,1,2,.... \qquad (1)$$

Thus $\alpha = \lambda t$ is the mean number of arrivals in time duration $[0, t]$

Now let's assume that the arrivals occur for the Poisson distributed arrival process as follows:

The first arrival occurs at time A_1, the second arrival occurs at time $A_1 + A_2$ and so on, assuming that the counting starts when time is 0. This is diagrammatically depicted as shown in the figure below:

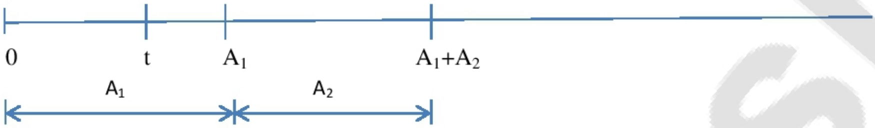

Fig (1) Arrival Process

Thus as can be seen from the Fig (1) above, $A_1, A_2, \ldots$ are successive inter-arrival times.

It is easy to observe that the first arrival occurs after time t if and only if there are no arrivals in the interval$[0, t]$.

Mathematically we express this as:

$$\{A_1 > t\} = \{N(t) = 0\}$$

and, therefore

$$P(A_1 > t) = P[N(t) = 0] = e^{-\lambda t} \qquad (2)$$

We get the above by substituting $n = 0$ in the Poisson distribution pdf in equation (1) above

Thus the probability that the first arrival will occur in interval $[0, t]$ is given as:

$$P(A_1 \leq t) = 1 - P(A_1 > t)$$

Substituting for $P(A_1 > t)$ from equation (2) in the above equation we get

$$P(A_1 \leq t) = 1 - e^{-\lambda t} \qquad (3)$$

But this is the cdf for an exponential distribution with parameter λ.

Hence we have shown that the first inter-arrival time $A_1$is distributed exponentially with mean $E(A_1) = 1/\lambda$. Similarly it can be shown that all the inter-arrival times A_1, A_2, A_3 .... are also exponentially distributed and independent with mean $1/\lambda$

Thus we have proved that "if arrival process follows Poisson distribution then inter-arrival time follows exponential distribution"

<u>Solution based upon the above concept for the problem of mail server</u>

Since the arrival of mails at a server follows Poisson Process with mean rate of arrivals λ, we know that the inter-arrival times between any two mail's arrival event must follow exponential distribution also with mean rate λ. (as proved above).

We have been given that mean rate of arrivals of the mails at server is one every 36 hours.

That is, $\lambda = \frac{1}{36}$

We want to find the probability that the next mail will occur between 24 and 48 hours after the last mail.

In effect we want to compute the probability of the event that the inter-arrival time of the next mail will be between 24 and 48 hours

Let A denote the random variable for the next inter-arrival time:

Since A is characterised by an exponential distribution with mean rate $= \frac{1}{36}$,

therefore:

$$\begin{aligned} P(24 \leq A \leq 48) &= F[48] - F[24] \\ &= \left(1 - e^{-\frac{48}{36}}\right) - \left(1 - e^{-\frac{24}{36}}\right) \\ &= e^{-2/3} - \ e^{-4/3} = 0.24981 \\ &\cong 0.25 \end{aligned}$$

Thus there is approximately 25% chance that the next mail will come between 24 and 48 hours after the last mail

*[**Guidelines to the Evaluator**: **i)** Even if the student has not formulated the proof for part 1 in this ideal step by step manner, if the basic method is logical and the analysis is broadly correct and sound full marks can be given for part 1. **ii)** For the second problem solving part of the question, if the method is correct but the answer is incorrect cut only 1 mark.*
***iii)** Allot 4 marks for the proof of part 1 and remaining 4 marks for the problem of part 2. Thus the marks distribution is equal for both the parts out of the total 8 marks for the question]*

Q. 5 (b) Random-variate generation of bus arrivals following poisson process by acceptance-rejection method using the given stream of 12 random numbers.

Solution: It is given that the buses arrive at the bus stop according to Poission process with mean of one bus per fifteen minutes.

Since 1 bus arrives per 15 minutes on the mean, therefore 4 buses should arrive per hour on the mean.

Therefore the parameter α for the Poisson process of arrivals per hour is 4

Since

$\propto = 4$, therefore, $e^{-\propto} = e^{-4} = 0.0183$ (1)

Now let's start the acceptance rejection technique for the generation of random variate N, representing the number of arriving buses during 1-hour time slot, according to Poisson process with mean $\propto = 4$. We would use the given sequence of 12 random numbers, as the source of consequent new random numbers required for this purpose.

Acceptance Rejection Technique of Random Variate Generation for Poisson Distribution:-

Step 1: Set $n = 0,\ P = 1$.

Step 2: $R_1 = 0.4357, P = 1. R_1 = 0.4357$

Step 3: Since $P = 0.4357 \geq e^{-\propto}$, reject $n = 0$ and return to Step 2, with $n = 1$.

Step 2: $R_2 = 0.4146, P = R_1. R_2 = 0.4357 \times 0.4146 = 0.1806$

Step 3: Since $P = 0.1806 \geq e^{-\propto}$, reject $n = 1$ and return to Step 2, with $n = 2$

Step 2: $R_3 = 0.8353, P = R_1. R_2. R_3 = 0.4357 \times 0.4146 \times 0.8353 = 0.1509$

Step 3: Since $P = 0.1509 \geq e^{-\propto}$, reject $n = 2$ and return to Step 2, with $n = 3$

Step 2: $R_4 = 0.9952, P = R_1. R_2. R_3. R_4 = 0.4357 \times 0.4146 \times 0.8353 \times 0.9952 = 0.1502$

Step 3: Since $P = 0.1502 \geq e^{-\propto}$, reject $n = 3$ and return to Step 2, with $n = 4$

Step 2: $R_5 = 0.8004, P = R_1. R_2. R_3. R_4. R_5 = 0.1202$

Step 3: Since $P = 0.1202 \geq e^{-\propto}$, reject $n = 4$ and return to Step 2, with $n = 5$

Step 2: $R_6 = 0.7945, P = R_1. R_2. R_3. R_4. R_5. R_6 = 0.0955$

Step 3: Since $P = 0.0955 \geq e^{-\propto}$, reject $n = 5$ and return to Step 2, with $n = 6$

Step 2: $R_7 = 0.1530, P = R_1. R_2. R_3. R_4. R_5. R_6. R_7 = 0.0146$

Step 3: Since $P = 0.0146 < e^{-\propto} = 0.0183$, accept $N = 6$.

Step 1: Set $n = 0, P = 1$

Step 2: $R_1 = 0.5678, P = 1. R_1 = 0.5678$

Step 3: Since $P = 0.5678 \geq e^{-\propto}$, reject $n = 0$ and return to Step 2 with $n = 1$

Step 2: $R_2 = 0.9854, P = R_1. R_2 = 0.5595$

Step 3: Since $P = 0.5595 \geq e^{-\propto}$, reject $n = 1$ and return to Step 2 with $n = 2$

Step 2: $R_3 = 0.1254, P = R_1. R_2. R_3 = 0.0702$

Step 3: Since $P = 0.0702 \geq e^{-\alpha}$, reject $n = 2$ and return to Step 2 with $n = 3$

Step 2: $R_4 = 0.6523, P = R_1. R_2. R_3. R_4 = 0.0458$

Step 3: Since $P = 0.0458 \geq e^{-\alpha}$, reject $n = 3$ and return to Step 2 with $n = 4$

Step 2: $R_5 = 0.2417, P = R_1. R_2. R_3. R_4. R_5 = 0.0111$

Step 3: Since $P = 0.0111 < e^{-\alpha} = 0.0183$, accept $N = 4$.

Thus we were able to generate 2 random-variate values for N; first we generated $N = 6$ and then we generated $N= 4$.

The point to be observed is that for accepting these 2 random-variate values we had to generate 12 values. Thus out of the total 12 values only 2 were finally accepted as values of N and the rest 10 had to be rejected.

*[**Guidelines to the Evaluator**: If the method is correct but the conclusion is incorrect you can cut 1 to 2 marks based on your discretion of whether the conclusion went wrong because of some silly mistake or because of deficient understanding on part of the student]*

Q.6 (a) Explain in detail "Time-Series input models."

Solution: Time-series Model is a model which is based on a sequence of related random variables.

Example	Variable	Alternate Model
Simulation of Web – based trading site of a stock broker	Time between arrivals of orders floated by investors to buy and sell. They are correlated to each other	Time Series Model with variables as time between arrivals of orders to buy and sell stocks.

Here, buy and sell orders tend to arrive in bursts, hence time between arrivals of orders to buy and sell stocks are dependent.

- **A time series is a sequence of random variables X_1, X_2, X_3, … , are identically distributed (same mean and variance) but dependent covariance-stationary.**
  - **cov(X_t, X_{t+h}) is the *lag-h auto-covariance***
  - **corr(X_t, X_{t+h}) is the *lag-h autocorrelation***
  - **If the auto-covariance values depends only on h and not on t, the time series is covariance-stationary**

There are two models that can be considered to study the relationship between the sequence of variables:

(1) AR(1) Model

(2) EAR(1) Model

as follows:

- **If X_1, X_2, X_3,… is a sequence of identically distributed, but dependent and covariance-stationary random variables, then we can represent the process as follows:**
  - **Autoregressive order-1 model, AR(1)**
  - **Exponential autoregressive order-1 model, EAR(1)**
    - **Both have the characteristics that:**
    - ***Lag-h* autocorrelation decreases geometrically as the lag increases, hence, observations far apart in time are nearly independent**

Q.6 (b) Finding the day of occurrence of the shortage condition after performing simulation of the inventory system assuming fixed initial inventory of 4 units.

Solution: The following table 1 shows the demand probabilities along with the intervals of random numbers that corresponds to each random daily demand.

Demand	Probability	Cumulative Probability	Ranges of random numbers
0	0.2	0.2	$(0, 0.2]$
1	0.5	0.7	$(0.2, 0.7]$
2	0.3	1.0	$(0.7, 1.0)$

The initial inventory before the simulation begins is given as 4

Let's perform the simulation for 5 days using the random numbers 0.4, 0.1, 0.8, 0.5 and 0.2 to generate daily demands for each day. The below table 2 shows the simulation logic for the 5 day inventory simulation:

Day	Start of Day inventory	Random No generated	Random Demand for the day	End of Day inventory	Did shortage occur on the day
1	4	0.4	1	$4 - 1 = 3$	No
2	3	0.1	0	$3 - 0 = 3$	No
3	3	0.8	2	$3 - 2 = 1$	No
4	1	0.5	1	$1 - 1 = 0$	No
5	0	0.2	0	$0 - 0 = 0$	No

From the above simulation table it is clear that the shortage condition does not occur on the first five days that were simulated. However, since at the end of the 5th day inventory level is zero, there is a chance that shortage might occur on any of the subsequent days if non zero demand is occurs on that day. Therefore shortage condition might occur on 6^{th} or subsequent days.

[***Important Note:*** *Note that on the fifth day the random number used is 0.2 which falls on the boundary of random number ranges marking cumulative probabilities. If some student considers 0.2 to fall in the range (0.2 to 0.7) instead of (0 to 0.2) for him demand of 1 will be generated instead of 0 on the fifth day. In such a case he might get a shortage condition on day 5 instead of no occurrence of shortage condition. Such a student according to me should also be considered for full marks if the rest of the method logic is correct*]

*[**Guidelines to the Evaluator**: **i)** If the method is correct but the conclusion is incorrect you can cut 1 to 2 marks based on your discretion of whether the conclusion went wrong because of some silly mistake or because of deficient understanding on part of the student*
*.**ii)** Both answers are plausible for this problem: 1) shortage condition does not occur & 2) shortage condition occurs on 5^{th} day depending on in which demand class the random number 0.2 is interpreted to fall. See **Important Note** given above]*

Q.7 (a) Write a short note on:- (i) Selecting input models without data

(ii) Validation process

- **Solution: (i) Selecting input models without data**

  Example: No funds or no time for preliminary study is available to collect the data before. In such a case, the following sources are used to obtain the information:

  – Engineering data: are the data values collected from the data published from performance ratings provided by the manufacturer or time specifications such as time for a process mentioned in company rules or production standards. These values provide the starting point for Input modeling.

  – Expert option(Educated Guess): are the data values generated through the people who are experienced with the process or similar processes, often, they can provide optimistic, pessimistic and most-likely times, and they may know the variability also.

  – Physical or conventional limitations: provides the boundary values or data due to the physical limits on performance, limits or bounds that narrow the range of the input process.

  E.g The space scarcity in South Mumbai may restrict the number of car arrivals at a service station.

The nature of the process:

  – In the absence of input data a certain distribution is assumed for the product or the process under consideration.

- The description of the known distributions can be used to fix the choice of a particular distribution.
- The uniform, triangular, and beta distributions are often used as input models.
- Here, the choice of Uniform is a poor choice as most of the time the lower 'a' and upper 'b' bounds of the assumed data are not known. Even if they are known they may be far away from the central value and hence may not give the correct idea of input data.

(ii) Validation process: See chapter 10 Jerry Banks 5th edition

Q.7 (b) Suppose five numbers generated are 0.44, 0.81, 0.14, 0.05, 0.93. Perform Kolmogorov-Smirnov test with level of significance α = 0.05, critical value D_α= 0.565.

Solution:

Step 1: Establishing null hypothesis H_0 and alternative hypothesis H_1

H_0: The generated numbers are uniformly distributed over the range space [0,1]

H_1: The generated numbers are NOT uniformly distributed over [0,1]

Step 2: Ranking of the random numbers from smallest to largest

After ranking the numbers from smallest to largest and labelling them as R_1,R_2, ….R_5 respectively we get

R_1 = 0.05, R_2= 0.14, R_3 = 0.44, R_4 = 0.81 and R_5 = 0.93

Step 3: Computing D^+ and D^-

D^+ is maximum of the set of all $+_{ve}$ deviations and D^- is the maximum of the set of all negative deviations.

Mathematically the set of all positive deviations and the set of all negative deviations is defined as:

+ve deviations = $\{i/N - R_i \mid 1 <= i <= N \;\&\; only\ positive\ values\ are\ to\ be\ considered\}$

-ve deviations = $\{i/N - R_i \mid 1 <= i <= N \;\&\; only\ positive\ values\ are\ to\ be\ considered\}$

The 4th and 5th rows of Table 2 below shows the set of all +ve deviations and the set of all

–ve deviations for our case:

i	1	2	3	4	5
R_i	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_i$	0.15	0.26	0.16	–	0.07
$R_i - (i-1)/N$	0.05	–	0.04	0.21	0.13

Thus we compute D^+ and D^- as:

$D^+ = max\,\{0.15, 0.26, 0.16, 0.07\} = 0.26$

$D^- = max\,\{0.05, 0.04, 0.21, 0.13\} = 0.21$

<u>Step 4: Computing the test statistic D = max (D^+, D^-)</u>

For our case we get

$D = max\,\{0.26, 0.21\} = 0.26$

<u>Step 5: Applying the test criteria using the pre-decided significance level α</u>

The significance level α and the critical value D_α is given to us as:

$\alpha = 0.05$ and $D_\alpha = 0.565$

Since the computed maximum deviation in step 4, $D = 0.26 < D_\alpha = 0.565$, the null hypothesis H_0, established in step 1, that the distribution of the generated numbers is uniform over [0,1] is not rejected

Conclusion: After performing the Kolmogorov-Smirnov test on the generated five numbers 0.44, 0.81, 0.14, 0.05 and 0.93 we conclude that no difference has been detected between the true distribution of { $R_1 = 0.05$, $R_2 = 0.14$, $R_3 = 0.44$, $R_4 = 0.81$ and $R_5 = 0.93$} and the ideal uniform distribution over [0,1].

*[**Note:** This is a solved example in Jerry Banks and John Carlsons book: Discrete Event System Simulation, 5th edition, See Example 6, Chapter 7: Random Number Generation under Section: Frequency Tests]*

*[**Guidelines to the Evaluator**: **i)** Even if the student has not formulated the solution in this ideal step by step manner of establishing the two hypothesis and then proceeding in the systematic manner full marks can be given if the basic method logic is correct and conclusion is also not wrong. **ii)** If the method is correct but the conclusion is incorrect you can cut 1 to 2 marks based on your discretion of whether the conclusion went wrong because of some silly mistake or because of deficient understanding on part of the student.]*