USN | | | | | | | | | |

CMRIT CMR INSTITUTE OF TECHNOLOGY, BENGALURU. ACCREDITED WITH A+ GRADE BY NAAC — CELEBRATING 25 YEARS

Internal Assessment Test II – April. 2018

Sub:	SYSTEM MODELING AND SIMULATION					Sub Code:	10CS82	Branch:	CSE & ISE
Date:	16-04-18	Duration:	90 min's	Max Marks:	50	Sem / Sec:	VIII/ A ,B,C(CSE) , VIII/A,B(ISE)		OBE

Answer any FIVE FULL Questions

		MARKS	CO	RBT
1 a.	Explain: i) Binomial Distribution ii)Uniform Distribution	[6]	CO3	L2

Sol:

i) Binomial Distribution – **3 Mark**

The no of successes in n Bernoulli trials is said to follow binomial distribution.

$P(x) = (n)\, p^{x} q^{n-x}$ for x=0,1,2,….n

Mean: E(x) = np

Variance : V(x) = npq

ii) Uniform Distribution – **3 Marks**

In probability theory and statistics, the **continuous uniform distribution** or **rectangular distribution** is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's support are equally probable. The support is defined by the two parameters, a and b, which are its minimum and maximum values.

Pdf : f(x) = 1/b-a , a<x<b

Cdf : F(x) = x-a/b-a , a<x<b

Mean: E(x) = a+b/2

Variance: $v(x) = (b-a)^2/12$

b.	Explain Discrete and continuous random variable with an example.	[4]	CO3	L2

Sol:

Discrete Random Variables ------2M

- *X* is a discrete random variable if the number of possible values of *X* is finite, or countably infinite.

Example: Consider jobs arriving at a job shop.

Let X be the number of jobs arriving each week at a job shop.

R_x= possible values of X (range space of X) = $\{0,1,2,\ldots\}$

$p(x_i)$ = probability the random variable is x_i is $P(X=x_i)$, $i = 1,2,$ ... must satisfy:

1. $p(x_i) \geq 0$, for all i

2. $\sum p(x_i) = 1$

The collection of pairs $[x_i, p(x_i)]$, $i = 1,2,\ldots,$ is called the probability distribution of X, and $p(x)$ is called the probability mass function (pmf) of X.

<u>Continuous Random Variables ------2M</u>

X is a continuous random variable if its range space R is an interval or a collection of intervals.

The probability that X lies in the interval $[a,b]$ is given by:

$$\int_{a}^{b} P(a = X = b) = f(x)dx$$

$f(x)$, denoted as the pdf of X, satisfies:

1. $f(x) = 0$, for all x in R

2. $f(x)dx = 1$

3. $f(x) = 0$, if x is not in R

Properties

1. $P(X = x_0 = 0) = 0$, because $x\int f(x)dx = 0$

2. $P(a = X = b) = P(a\, X = b) = P(a = X\, b) = P(a\, X\, b)$

Example: Life of an inspection device is given by X, a continuous random variable with pdf:

$f(x)=1/2e^{-x}$,x=0

=0,otherwise

X has an exponential distribution with mean 2 years

- Probability that the device's life is between 2 and 3 years is:

  $P(2<=x<=3)= 1/2\int_{2}^{3} e^{-x/2}\, dx = 0.14$

2 a. What are Pseudo random numbers? What are the problems that occur while generating pseudo random numbers? Also list the important considerations during generation of pseudo random numbers. [5] CO4 L1

Sol:

Random NO generation

→ properties of random no's

→ uniformity (0 and 1)

→ Independent

→ Generation of psedo random no's

1) what are psedo random numbers what are the problems that occur while generating psedo random no? Also list the imp considerations during generat of psedo random no's?

'psedo' means false, so false numbers are being generated.

→ psedo implies act of generating random no by a known method removed the true randomness

→ certain problems (or) errors occur in the generation of psedo random numbers are

→ the generated no's might not be uniformly distributed

→ the generated no's might be discrete valued instead of continuous valued.

→ the mean of the generated no's might be too high (or) low

→ the variance might be too high (or) too low

→ there might be dependence

→ Random no's are generated by a digital computer as part of the simulation. the important consideration for random no generation methods are as follows

→ the generated random no should follow important properties of uniformity and independence

→ the routine should be portable to different computer languages.

→ the routine have a sufficiently long cycle.

→ The routine should be fast.

→ the random no should be replicable. Given a starting point it should be possible to generate same set of random nos.

b. Using suitable frequency test find out whether the random numbers generated are uniformly distributed on the interval [0, 1] can be rejected. Assume α=0.05 and Dα=0.565.the random numbers are 0.54, 0.73, 0.98, 0.11, 0.68 [5] CO4 L3

Sol:

- Rank the data from smallest to largest.Finding the R_i - **1 Mark**
  - i.e. R_i 0.11 0.54 0.68 0.73 0.98

- Finding the D^+ and D^- values – **2 Marks**
  - i.e. $D^+ = i/N - R_i = 0.09$
  - $D^- = R_i - (i-1)/N = 0.34$
- Finding D value – **1 Mark**
  - i.e. $D = \max(D^+, D^-) = 0.34$
- Justification of the data accept or reject – **1 Marks**
  - i.e. The tabular value $D_{\alpha,n} = 0.565$.Since $D < D_\alpha$ i.e $0.34 < 0.565$ the sequence of numbers given are accepted.

3	Explain acceptance –rejection technique? Generate five Poisson variates with mean 0.25 by taking these random numbers 0.073, 0.693, 0.945, 0.739, 0.014, 0.342	[10]	CO4	L3

Sol:

Suppose that an analyst needed to devise a method for generating random variates, X, uniformly distributed between 0 and 1. One way to proceed would be to follow these steps:

Step1: Generate a random number R.

Step 2a. If R > 1/4, accept X = /?, then go to step 3.

Step 2b. If R < 1/4, reject /?, and return to step 1.

Step 3. If another uniform random variate on [1/4,1] is needed, repeat the procedure beginning at step 1. If not, stop.

Each time step 1 is executed, a new random number R must be generated. Step 2a is an "acceptance" and step 2b is a "rejection" in this acceptance rejection technique. To summarize the technique, random variates (R) with some distribution (here uniform on [0,1]) are generated until some condition (R > 1/4) is satisfied. When the condition is finally satisfied, the desired random variate, X (here uniform on [1/4,1]), can be computed (X = R). This procedure can be shown to be correct by recognizing that the accepted values of R are conditioned values; that is, R itself does not have the desired distribution, but R conditioned on the event {R > 1/4} does have the desired distribution.

To show this, take 1/4 < a < b < 1; then

$$P (a < R \leq b \mid . \leq R \leq 1) = P (a < R \leq b) / P (. \leq R \leq 1) = b - a / ¾$$

which is the correct probability for a uniform distribution on [1/4, 1]. Equation (5.28) says that the probability distribution of /?, given that R is between 1/4 and 1 (all other values of R are thrown out), is the desired distribution. Therefore, if 1/4 < R < 1, set X = R.

- Explanation of acceptance rejection technique – **2 Marks**
- Finding five poisson variates – **8 Marks**
  - i.e. 1. Set n=0,p=1.
  - 2. R1=0.073,P=1*0.073 = 0.073
  - 3. Since $P=0.073 < e^{-0.25} = 0.7788$, accept N=0 - **2 Marks**
  - Steps 1-3.(R1=0.693 leads to N=0),accept N=0 – **2 Marks**
  - 1.Set n=0,P=1

- 2. R1=0.945,P=1*0.945=0.945
- 3.Since 0.945 >0.7788,reject n=0 and return to step 2 with n=1
- Step 2. R2= 0.739,P=R1R2=0.945*0.739 = 0.698
- 3. Since 0.698 < 0.7788 accept N=1 – 1 **Marks**
- Steps 1-3.(R1=0.014 leads to N=0),accept N=0 – 1 **Marks**
- Steps 1-3.(R1=0.342 leads to N=0),accept N=0 – **2 Marks**
- The five Poisson Variates are 0.073,0.693,0.739,0.014,0.342
-

n	R_{n+1}	P	Accept/Reject	Result
0	0.073	0.073	$P < e^{-b}$ (Accept)	N=0
0	0.693	0.693	$P < e^{-b}$ (Accept)	N=0
0	0.945	0.945	$P > e^{-b}$ (Reject)	
1	0.739	0.698	$P < e^{-b}$ (Accept)	N=1
0	0.014	0.014	$P < e^{-b}$ (Accept)	N=0
0	0.342	0.342	$P < e^{-b}$ (Accept)	N=0

[10]

CO4 L3

4 Consider the following 60 values. Test whether 2^{nd}, 9^{th}, 16^{th},……,numbers in the sequence are auto correlated for α=0.05. (Consider the tabular value $Z_{0.025} = 1.96$).

0.30	0.48	0.36	0.01	0.54	0.34	0.96	0.06	0.61	0.85
0.48	0.86	0.14	0.86	0.89	0.37	0.49	0.60	0.04	0.83
0.42	0.83	0.37	0.21	0.90	0.89	0.91	0.79	0.57	0.99
0.95	0.27	0.41	0.81	0.96	0.31	0.09	0.06	0.23	0.77
0.73	0.72	0.47	0.13	0.55	0.11	0.75	0.36	0.25	0.23
0.64	0.84	0.76	0.30	0.26	0.38	0.05	0.19	0.73	0.44

Sol :

- ➢ Finding the autocorrelation value along with steps– **10 Marks**

The marks split up are as follows.

- Finding the M value – **2 Mark**

  i.e. i+(M+1)l <= N

- Here i=2, l= 7 and N=60
-

2+(M+1)7 <= 60.For M=8 this condition won't satisfy so the previous value is M=7.

- Finding the ρ_{im} value – **4 Marks** i.e. ρ_{im} = 1/M+1[R2.R9+R9.R16+R16.R23+R23.R30+R30.R37+R37.R44+R44.R51+R51.R58] - 0.25

=1/8[(0.48)(0.61)+(0.61)(0.37)+(0.37)(0.37)+(0.37)(0.99)+(0.99)(0.09)+(0.09)(0.55)+(0.55)(0.60)+(0.60)(0.19)] – 0.25

=1/8(1.6043) – 0.25 = -0.494

- Finding the $\sigma_{\rho im}$ = ▪ 13M+7 / 12(M+1) value – **1 Mark**

  =▪ 13*7+7 / 12(8) = 0.1031

- Finding Z_0 value - **1 Mark**

  - i.e. Z_0 = -0.0494/0.1031 = -0.4791

➢ Justification of acceptance – **2 Marks**

i.e. $-Z_{\alpha/2} <= Z_0 <= Z_{\alpha/2}$
= -1.96 < -0.4791 <1.96.Hence the sequence of numbers given are accepted

5. Use the Chi-square test with α=0.05, n=10, $x^2_{0.05.9}$ =16.9 to test for whether the data shown are uniformly distributed.	CO4	L3

0.34	0.9	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.7	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.3	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.1	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.3	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.4	0.64	0.4	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.6	0.11	0.29	0.78

Sol:

➢ Dividing into no of intervals-**1 Mark**
➢ Finding Ei value – **1 Mark**
   Ei= N/n = 100/10 =10
➢ Chi-Square test computation table – **7 Marks**

Interval No	Interval range	Oi	Ei	Oi-Ei	(Oi-Ei)²	(Oi-Ei)²/Ei
1	0-0.1	8	10	-2	4	0.4
2	0.1-0.2	8	10	-2	4	0.4
3	0.2-0.3	10	10	0	0	0
4	0.3-0.4	9	10	-1	1	0.1
5	0.4-0.5	12	10	2	4	0.4
6	0.5-0.6	8	10	-2	4	0.4
7	0.6-0.7	10	10	0	0	0
8	0.7-.0.8	14	10	4	16	1.6
9	0.8-0.9	10	10	0	0	0
10	0.9-1.0	11	10	1	1	0.1
Total						3.4

Comparing with α value and justifying for acceptance/rejection – **1 Mark**
i.e.3.4 < 16.9 so the above values are accepted

6. Explain Chi–square goodness of fit test. Apply it to Poisson assumption with α=3.64.Data size=100 and observed frequency Oi=12,10,19,17,10,8,7,5,5,3,3,1

Sol:

For Poisson distribution $P(x) = e^{-\alpha}\alpha^{x} / x!$ for x=0,1,2,…..

Compute P(0),P(1),P(2),……P(11) as follows – **2 Marks**

$P(0)= e^{-3.64}(3.64)^{0} / 0! = 0.026$
$P(1)= e^{-3.64}(3.64)^{1} / 1! = 0.096$
$P(2)= e^{-3.64}(3.64)^{2} / 2! = 0.174$
$P(3)= e^{-3.64}(3.64)^{3} / 3! = 0.211$
……till P(11) as follows

- Chi-square test Table- **7 Marks**

[10] CO5 L3

$P(4) = \frac{e^{-3.64}(3.64)^4}{4!} = 0.192$

$P(5) = \frac{e^{-3.64}(3.64)^5}{5!} = 0.140$

$P(6) = \frac{e^{-3.64}(3.64)^6}{6!} = 0.085$

$P(7) = \frac{e^{-3.64}(3.64)^7}{7!} = 0.044$

$P(8) = \frac{e^{-3.64}(3.64)^8}{8!} = 0.020$

$P(9) = \frac{e^{-3.64}(3.64)^9}{9!} = 0.008$

$P(10) = \frac{e^{-3.64}(3.64)^{10}}{10!} = 0.003$

$P(11) = \frac{e^{-3.64}(3.64)^{11}}{11!} = 0.001$

chi-square test table

x_i	observed frequency O_i		Expected frequency E_i ($n \times p_i$)		$\frac{(O_i-E_i)^2}{E_i}$
0	12	22	$100 \times 0.026 = 2.6$	12.2	7.87
1	10		$100 \times 0.096 = 9.6$		
2	19		$100 \times 0.174 = 17.4$		0.15
3	17		$100 \times 0.211 = 21.1$		0.80
4	10		$100 \times 0.192 = 19.2$		4.41
5	8		$100 \times 0.140 = 14.0$		2.57
6	7		$100 \times 0.085 = 8.5$		0.26
7	5	17	$100 \times 0.044 = 4.4$	7.6	11.62
8	5		$100 \times 0.020 = 2.0$		
9	3		$100 \times 0.008 = 0.8$		
10	3		$100 \times 0.003 = 0.3$		
11	1		$100 \times 0.001 = 0.1$		
					27.68

— Always we have to see that the expected freq values should be >5. If not then combine with previous (or) next value until it becomes >5.

7 Explain different steps in the development of a useful model of input data [10] CO5 L2

Sol:

There are **four steps** in the development of a useful model of input data

• *Collect data from the real system of interest.* This often requires a substantial time and resource commitment. Unfortunately, in some situations it is not possible to collect data.

• *Identify a probability distribution to represent the input process*. When data are available, this step typically begins by developing a frequency distribution, or histogram, of the data.

• *Choose parameters that determine a specific instance of the distribution family*. When data are available, these parameters may be estimated from the data.

• *Evaluate the chosen distribution and the associated parameters for good-of-fit*. Goodness-of-fit may be evaluated informally via graphical methods, or formally via Statistical tests.

1)Data Collection:

The following suggestions may enhance and facilitate data collection, although they are not all – inclusive.

1. *A useful expenditure of time is in planning*. This could begin by a practice or pre observing session. Try to collect data while pre observing.

2. *Try to analyze the data as they are being collected*. Determine if any data being collected are useless to the simulation. There is no need to collect superfluous data.

3. *Try to combine homogeneous data sets*. Check data for homogeneity in successive time periods and during the same time period on successive days.

4. *Be aware of the possibility of data censoring, in which a quantity of interest is not observed in its entirety.* This problem most often occurs when the analyst is interested in the time required to complete some process (for example, produce a part, treat a patient, or have a component fail), but the process begins prior to, or finishes after the completion of, the observation period.

5. *To determine whether there is a relationship between two variables, build a scatter diagram.*

6. *Consider the possibility that a sequence of observations which appear to be independent may possess autocorrelation*. Autocorrelation may exist in successive time periods or for successive customers.

7. *Keep in mind the difference between input data and output or performance data, and be sure to collect input data*. Input data typically represent the uncertain quantities that are largely beyond the control of the system and will not be altered by changes made to improve the system.

2)Histogram

A frequency distribution or histogram is useful in identifying the shape of a distribution.

A histogram is constructed as follows:

1. Divide the range of the data into intervals (intervals are usually of equal width; however, Unequal widths however, unequal width may be used if the heights of the frequencies are adjusted).

2. Label the horizontal axis to conform to the intervals selected.
3. Determine the frequency of occurrences within each interval.
4. Label the vertical axis so that the total occurrences can be plotted for each interval.
5. Plot the frequencies on the vertical axis.

- If the intervals are too wide, the histogram will be coarse, or blocky, and its shape and other details will not show well. If the intervals are too narrow, the histogram will be ragged and will not smooth the data.
- The histogram for continuous data corresponds to the probability density function of a theoretical distribution.

<u>3) Selecting the Family of Distributions</u>

Additionally,the shapes of these distributions were displayed. The purpose of preparing histogram is to infer a known pdf or pmf. A family of distributions is selected on the basis of what might arise in the context being investigated along with the shape of the histogram.

Thus, if inter arrival-time data have been collected and the histogram has a shape similar to the pdf in Figure 5.9.the assumption of an exponential distribution would be warranted.

- Similarly, if measurements of weights of pallets of freight are being made, and the histogram appears symmetric about the mean with a shape like that shown in Fig 5.12, the assumption of a normal distribution would be warranted.
- The exponential, normal, and Poisson distributions are frequently encountered And are not difficult to analyze from a computational standpoint.
  Although more difficult to analyze, the gamma and Weibull distributions provide array of shapes, and should not be overlooked when modeling an underlying probabilistic process. Perhaps an exponential distribution was assumed, but it was found not to fit the data. The next step would be to examine where the lack of fit occurred.
- If the lack of fit was in one of the tails of the distribution, perhaps a gamma or Weibull distribution would more adequately fit the data.
- Literally hundreds of probability distributions have been created, many with Some specific physical process in mind. One aid to selecting distributions is to use the physical basis of the distributions as a guide. Here are some examples:

Binomial: Models the number of successes in n trials, when the trials are independent With common success probability, p; for example, the number of defective computer chips found in a lot of n chips.

Negative Binomial (includes the geometric distribution) : Models the number of trials required to achieve k successes; for example, the number of computer chips that we must inspect to find 4 defective chips.

Poisson: Models the number of independent events that occur in a fixed amount of

time or space: for example, the number of customers that arrive to a store during 1 hour, or the number of defects found in 30 square meters of sheet metal.
Normal: Models the distribution of a process that can be thought of as the sum of a number of component processes; for example, the time to assemble a product which is the sum of the times required for each assembly operation. Notice that the normal distribution admits negative values, which may be-impossible for process times.

Lognormal: Models the distribution of a process that can be thought of as the product of (meaning to multiply together) a number of component processes; for example, the rate of return on an investment, when interest is compounded, is the product of the returns for a number of periods.

Exponential: Models the time between independent events, or a process time which is Memory less (knowing how much time has passed gives no information about how Much additional time will pass before the process is complete); for example, the times between the arrivals of a large number of customers who act independently of each other.

Gamma: An extremely flexible distribution used to model nonnegative random variables. The gamma can be shifted away from 0 by adding a constant.

Beta: An extremely flexible distribution used to model bounded (fixed upper and lower limits) random variables. The beta can be shifted away from 0 by adding a constant and can have a larger range than [0,1] by multiplying by a constant.

Weibull: Models the time to failure for components; for example, the time to failure for a disk drive. The exponential is a special case of the Weibull. Discrete or Continuous Uniform Models complete uncertainty, since all outcomes are equally likely. This distribution is often overused when there are no data.

Triangular Models a process when only the minimum, most-likely, and maximum values of the distribution are known; for example, the minimum, most-likely and maximum time required testing a product.

Empirical Re-samples from the actual data collected; often used when no theoretical distribution seems appropriate.

4)Parameter Estimation

After a family of distributions has been selected, the next step is to estimate the parameters of the distribution.

Suggested Estimators: Numerical estimates of the distribution parameters are needed to Reduce the family of distributions to a specific distribution and to test the resulting hypothesis.
These estimators are the maximum-likelihood estimators based on the raw data. (If the data are in class intervals, these estimators must be modified.). The triangular distribution is usually employed when no data are available; with the parameters obtained from educated guesses for the minimum, most likely, and maximum possible values the uniform distribution may also be used in this way if only minimum and maximum values are available.
Examples of the use of the estimators are given in the following paragraphs. The

reader should keep in mind that a parameter is an unknown constant, but the estimator is a statistic or random variable because it depends on the sample values.

8 The time required for 50 different employees to compute and record the number of hours during the week was measured with the following results in minutes. Use goodness of fittest to test the hypothesis that these service times are exponentially distributed. Assume k=6, λ=1.206, α = 0.05.Use $\chi^2_{0.05, 4} = 9.49$. [10] CO5 L4

1.88	1.54	1.90	0.15	0.02	2.81	1.50	0.53	2.62	2.67
3.53	0.53	1.80	0.79	0.21	0.80	0.26	0.63	0.36	2.03
1.42	1.28	0.82	2.16	0.05	0.04	1.49	0.66	2.03	1.00
0.39	0.34	0.01	0.10	1.10	0.24	0.26	0.45	0.17	4.29
0.80	5.50	4.91	0.35	0.36	0.90	1.03	1.73	0.38	0.48

Sol:

- Finding P value – **1Mark**
  - P=1/k = 1/6 = 0.1667
- Finding Expected Values – **1 Mark**
  - Ei=nPi=50*0.1667 = 8.33
- Finding A0,A1,….values – **2 Marks**

  $A_i = -1/\lambda \ln(1-iP)$

  A0=-1/1.206(1-0*0.1667)=0

  A1=-1/1.206 ln(1-1*0.1667) = 0.1512

  A2= -1/1.206 ln(1-2*0.1667) = 0.3362

  A3= -1/1.206 ln(1-3*0.1667) = 0.5749

  A4= -1/1.206 ln(1-4*0.1667) = 0.9112

  A5 = -1/1.206 ln(1-5*0.1677) = 1.4865

- Chi-square table – **5 Marks**

Interval	**Oi**	**Ei**	**Oi-Ei**	**(Oi-Ei)2**	**(Oi-Ei)2/Ei**
0-0.1512	6	8.33	2.33	5.428	0.6517
0.1512 - 0.3362	6	8.33	2.33	5.428	0.6517
0.3362 - 0.5749	10	8.33	1.67	2.788	0.334
0.5749 - 0.9112	7	8.33	1.33	1.768	0.2123
0.9112 - 1.4865	5	8.33	3.33	11.088	1.331
1.4865 - ∞	16	8.33	7.67	58.82	7.0622
				Total	**10.24**

- Justification of acceptance – **1 Mark**
  - 10.24 > 9.49 so the data given are rejected.