ACE-DS: Modelling and Simulation module, Preparation for the Exam

Notes: The exam will be a **FULLY OPEN BOOK** exam.

Question 1. Tankers arrive at an oil port to be unloaded with the following distribution of interarrival times:

Interarrival times (in days)	Probability
1	0.35
2	0.25
3	0.25
4	0.10
5	0.05

The port has two terminals, A and B. The time it takes to unload a tanker is a random variable with the following distribution (both terminals A and B have the same unloading times):

time (in days)	Probability
2	0.15
3	0.35
4	0.50

Arriving tankers form a single waiting line in the port area until a terminal becomes available for service. Service is given on first-come-first-serviced basis.
Develop a simulation model to simulate the arrival and unloading of oil tankers in the port in order to determine the average time a tank spends in the port and the time-average number of tanks in the port. Use ARENA to run your model for 360 days to determine these performance measures. Note that in ARENA, the discrete distribution taking values on the set $\{V_1, V_2, \dots, V_n\}$ is entered as: `EXPRESSION` and the expression is:

$$\texttt{DISC}(P_1, V_1, P_2, V_2, \dots, P_n, V_n)$$

where $P_1, P_2, \dots, P_n$ are the **cumulative probabilities**. That is:

$$P_1 = p(V_1), P2 = p(V_1) + p(V_2), \dots, Pn = p(V_1) + p(V_2) + \dots + p(V_n) = 1.$$

[10 MARKS]

Question 2. Travelers arrive at the main entrance door of an airline terminal according to an exponential interarrival-time distribution with mean 1.6 minutes, with the first arrival at time 0. The travel time from the entrance to the check-in is distributed uniformly between 2 and 3 minutes. At the check-in counter, travelers wait in a single line until one of five agents is available to serve them. The check-in time (in minutes) follows a Weibull distribution with parameters $\beta = 7.78$ and $\alpha = 3.91$. (Note that in Arena this distribution is denoted WEIB(β, α).) Upon completion of their check-in, they are free to travel to their gates. Create a simulation model of this system. Run the simulation for a single replication of 16 hours to determine the average time in system, number of passengers completing check-in, and the time-average length of the check-in queue.

[10 MARKS]

Question 3 Develop a model of a simple serial two-process system. Items arrive at the system with a mean time between arrivals of 10 minutes, with the first arrival at time 0. They are immediately sent to Process 1, which has a single resource with a mean service time of 9.1 minutes. Upon completion, they are sent to Process 2, which is identical to (but independent of) Process 1. Items depart the system upon completion of Process 2. Performance measures of interest are the time-average numbers in queue at each process and the average total time in system of items. Using a single replication of length 10,000 minutes, make the following four runs and compare the results (noting that the structure of the model is unchanged, and it's only the input distributions that are changing):

```
Run 1: exponential interarrival times and exponential service times

Run 2: constant interarrival times and exponential service times

Run 3: exponential interarrival times and constant service times

Run 4: constant interarrival times and constant service times
```

[10 MARKS]

Question 4 Parts arrive at a single workstation system according to an exponential interarrival distribution with mean 21.5 seconds; the first arrival is at time 0. Upon arrival, the parts are initially processed. The processing-time distribution is TRIA(16, 19, 22) seconds. There are several easily identifiable visual characteristics that determine whether a part has a potential quality problem. These parts, about 10% (determined after the initial processing), are sent to a station where they undergo a thorough inspection. The remaining parts are considered good and are sent out of the system. The inspection-time distribution is 95 plus a WEIB(48.5, 4.04) random variable, in seconds. About 14% of these parts fail the inspection and are sent to scrap. The parts that pass the inspection are classified as good and are sent out of the system (so these parts didn't need the thorough inspection, but you know what they say about hindsight). Run the simulation for 10,000 seconds to observe the number of good parts that exit the system, the number of scrapped parts, and the number of parts that received the thorough inspection.

[10 MARKS]

Question 5. The Rayleigh distribution of parameter σ is such that its cumulative distribution function (CDF) is given by:

$$F(x) = \begin{cases} 1 - e^{-x^2/2\sigma^2} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Apply the *inverse transform method* to show that the following algorithm generates random numbers from the Rayleigh distribution of parameter σ:

(Step 1) Generate U from $U[0, 1)$;

(Step 2) Return $X = \sqrt{-2\sigma^2 \ln(1 - U)}$.

(b) Write a R code of the algorithm given in question (a) to generate 1000 random numbers from the Rayleigh distribution with parameter $\sigma = 3$. Give the histogram of your output.

(c) It is well-known that if Y is a random variable with the Rayleigh distribution with parameter σ, then Y^2 has the exponential distribution with rate $\lambda = \frac{1}{2\sigma^2}$. Now use the following R code to sample $n = 1000$ random numbers from the random variable Y with Rayleigh distribution of parameter $\sigma = 3$:

```
library(extraDistr)
rrayleigh(n, sigma)
```

(You may need to install the package "extraDistr", this will take a few seconds.) Deduce a sample of n random numbers from the random variable Y^2. Finally perform goodness-of-fit tests to test whether indeed your output (for Y^2) comes from the exponential distribution of rate $\frac{1}{18}$. Comment on your result.

[10 MARKS]

Question 6. Over time, bonds are liable to move from one rating category to another. This is sometimes referred to as *credit ratings migration*. Rating agencies produce from historical data a rating transition matrix. This matrix shows the probabiltiy of a bond moving from one rating to another during a certain period of time. Usually the period of time is one year. Below is a table giving a rating transition matrix produced from historical data by Standard and Poor's (S&P)

Table 1: Table 1: One-year transition probabilities matrix

	Ratings at year-end							
Initial ratings	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.9366	0.0583	0.0040	0.0009	0.0002	0	0	0
AA	0.0066	0.9172	0.0694	0.0049	0.0006	0.0009	0.0002	0.0002
A	0.0007	0.0225	0.9176	0.0518	0.0049	0.0020	0.0001	0.0004
BBB	0.0003	0.0026	0.0483	0.8924	0.0444	0.0081	0.0016	0.0023
BB	0.0003	0.0006	0.0044	0.0666	0.8323	0.0746	0.0105	0.0107
B	0	0.0010	0.0032	0.0046	0.0572	0.8362	0.0384	0.0594
CCC	0.0015	0	0.0029	0.0088	0.0191	0.1028	0.6123	0.2526
Default	0	0	0	0	0	0	0	1.0000

(Source: Standard & Poor's, January 2001)

For ease of reference we shall denote the states or ratings `AAA, AA, A, BBB, BB, B, CCC, Default` as states `1, 2, 3, 4, 5, 6, 7, 8` respectively. That is AAA is state 1, AA state 2, etc.

In reality, this transition matrix is updated every year. However if assuming no significant change in the transition matrix in the future, then one can use the transition matrix to predict what will happen over several years in the future. In particular, one can regard the transition matrix as a specification of a Markov chain model with states `1, 2, 3, 4, 5, 6, 7, 8`.

Because it might take you some time to copy the transition matrix in R, it is rewritten here in a form ready to be used in R so that you can simply copy and paste where necessary.

```
P <- matrix(c(0.9366, 0.0583, 0.0040, 0.0009, 0.0002, 0,      0,      0,
              0.0066, 0.9172, 0.0694, 0.0049, 0.0006, 0.0009, 0.0002, 0.000,
              0.0007, 0.0225, 0.9176, 0.0518, 0.0049, 0.0020, 0.0001, 0.0004,
              0.0003, 0.0026, 0.0483, 0.8924, 0.0444, 0.0081, 0.0016, 0.0023,
              0.0003, 0.0006, 0.0044, 0.0666, 0.8323, 0.0746, 0.0105, 0.0107,
              0,      0.0010, 0.0032, 0.0046, 0.0572, 0.8362, 0.0384, 0.0594,
              0.0015, 0,      0.0029, 0.0088, 0.0191, 0.1028, 0.6123, 0.2526,
              0,      0,      0,      0,      0,      0,      0,      1.000),
                 nrow = 8, ncol = 8, byrow = TRUE)
```

Answer the following questions:

(a) What is the probability that a currently "AA" rated bond becomes "BB" after 5 years?

(b) Assume that a given bond is currently in rating "AA" (that is state 2). Write a program which generates a random path $X_1, X_2, \dots, X_N$ from this Markov chain with $N = 1000$. (Use the fact that X_0 is "AA" or $X_0 = 2$.)

(c) Consider your output in subquestion (b). Did your chain the reach state "AAA"? Substiante your answer.

[10 MARKS]

Question 7. Based on statistical theory, it is known that if one takes take a sample of size $n = 15$ from a $N(0, 1)$ distribution, then

$$t = \frac{\bar{X}}{s/\sqrt{15}}$$

follows a t distribution with $n - 1 = 14$ degrees of freedom. Here $\bar{X}$ is the sample mean and s is the sample standard deviation given by

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

and

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_i - \bar{X})^2.$$

(a) Use simulation to confirm this theoretical result: Generate 1000 random values of t using the given formula, give the corresponding histogram and overlay the density of the t distribution.

(b) Finally apply the Kolmogorov-Smirnov goodness-of-fit test to test whether indeed your sample is from the t distribution with degree of freedom 14. Note that in R to generate n random numbers from the t distribution with degree of freedom `df`, one can simply use `rt(n, df)`. Moreover the probability density function of the t distribution with degree of freedom `df` is given by `dt(x, df)`.

[10 MARKS]

Question 8. Determine which of the following mixed linear congruential generators have a full period. Justify all your claims.

(a) $Z_i = (13Z_{i-1} + 13) \mod 16, \; Z_0 = 1$

(b) $Z_i = (12Z_{i-1} + 13) \mod 16, \; Z_0 = 2$

(c) $Z_i = (13Z_{i-1} + 12) \mod 16, \; Z_0 = 1$

(d) $Z_i = \quad (Z_{i-1} + 12) \mod 13, \; Z_0 = 3$

[10 MARKS]

Question 9. Let $X = (X_1, X_2, X_3, X_4, X_5, X_6)$ be a multivariate random variable. Assume that it has the multivariate normal distribution $N(\mu, \Sigma)$ in $\mathbb{R}^6$ where $\mu = (2, 4, -2, 1, 3, -2)$ and Σ is the matrix given by

$$\Sigma = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

Generate 1000 random samples from this distribution and use your sample to estimate the probability that

$$(X_1 + X_2 + X_3) < 2(X_4 + X_5 + X_6).$$

[10 MARKS]

Question 10. In modelling a casualty department of a hospital, the following problem is identified. The processing of patients requires four stages:

(i) *triage*, where each patient's obvious symptoms are identified,

(ii) *initial treatment*, where minor wounds are treated or first aid is administrated for major problems,

(iii) *X-ray*, where detailed checks are made on serious trauma cases, and

(iv) *final treatment*, where serious problems are handled before sending them home or admitting them to a hospital ward.

After the stages of initial treatment and final treatment some patients are sent home. Not all patients require X-rays.

The personnel are overwhelmed with very long queues and patients are complaining about waiting times. As an analyst, you are asked to build a simulation model for this department of the hospital. Answer the following questions:

- What do you think should be the objective of the study?
- Which simulation model is appropriate to the system: discrete or continuous, static or dynamic, deterministic or stochastic? Justify your claims.
- Build a sketchy simulation model for this system. Present it in a form of a (single) flowchart.
- It is clear that no data are provided. Explain in detail, which data are relevant in order to build a meaningful simulation model for this system and how they can be obtained.
- Briefly explain how the simulation model can be used by the management to improve service delivery in the hospital.

[10 MARKS]

Question 11. Use simulation to estimate the value of the integral:

$$\int_0^3 e^{-x^2/2} dx.$$

Consider 10000 iterations.

[10 MARKS]

Question 12. Use simulation to estimate the expected value of the quantity

$$f(X) = e^{2X+3}$$

where X is a random variable following the standard normal distribution.

[10 MARKS]

Question 13. Given a sequence $X_1, X_2, \dots,$ of $U[0,1]$-distributed pseudo random numbers, we can use a scatter plot of (X_i, X_{i+1}) for $i = 1, ..., n-1$ in order to try to assess whether the X_i are independent.

(a) Create such a plot using the built-in random number generator of R:

```
X <- runif(1000)
plot(X[1:999], X[2:1000], asp=1)
```

Can you explain the resulting plot?

(b) Create a similar plot, using your function LCG from Question 7

```
m <- 81
a <- 1
c <- 8
seed <- 0
X <- LCG(1000, m, a, c, seed)/m
plot(X[1:999], X[2:1000], asp=1)
```

Discuss the resulting plot.

(c) Repeat the experiment from (b) using the parameters

$$m = 1024, a = 401, c = 101$$

and

$$m = 2^{32}, a = 1664525, c = 1013904223$$

. Discuss the results.

[10 MARKS]

============== END==============