Module 3E3

RISK MODELLING

Crib

QUESTION 1.

1(a) **(i)** Average service time: $1/\mu=1/30$ hours= 2 min.
Expected service completion time= 2:05pm + 0:02= 2:00 pm

(ii) Average time between arrivals: $1/\lambda=1/24=2.5$ min.
Expected time of the next arrival: 4:00:00 pm + 00:02:30 = 4:02:30 pm

(iii) % of time busy= Average utilisation rate when s=1: $\lambda/(s\mu)=24/(1\times30)=0.8$
Therefore, 80% of time the ATM is busy.

(iv) Average number waiting in line: $L_q=\lambda.W_q=24\times 0.1333= 3.1992 \cong 3.2$ students

(v) P(wait)= P (All ATMs are busy)= 0.8.

(vi) The total hourly cost of customers' delays in queues = λ W_q $\times$ (cost per hour per student)= $24\times0.1333\times$ \$10/ hour=\$32 per hour

(vii) s=2; Average utilisation rate when s=2: $\lambda/(s\mu)=24/(2\times30)=0.4$= 40%.

(viii) s=2; P(student will not wait)= $P_0 + P_1$

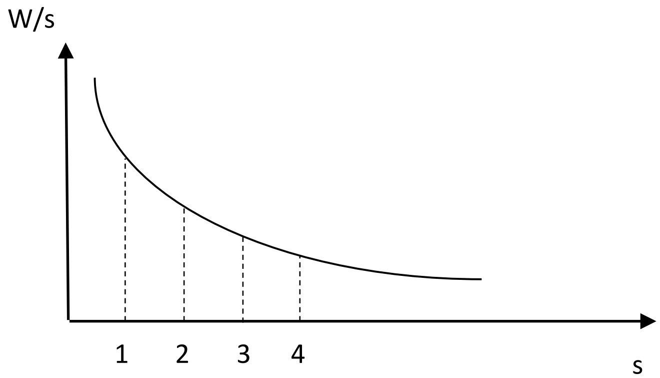
$$p_0=\frac{1}{\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!}+\frac{(\lambda/\mu)^s}{s!}\left(\frac{s\mu}{s\mu-\lambda}\right)}$$

$$p_n=\begin{cases}\frac{(\lambda/\mu)^n}{n!}p_0 & if\ 0\le n\le s\\ \frac{(\lambda/\mu)^n}{s!s^{n-s}}p_0 & if\ n\ge s\end{cases}$$

$$L_q=\left(\frac{(\lambda/\mu)^{s+1}}{(s-1)!(s-\lambda/\mu)^2}\right)p_0.$$

$$P_0=\frac{1}{1+{}^{24}/_{30}+\frac{\left({}^{24}/_{30}\right)^2}{2!}\left(\frac{2*30}{2*30-24}\right)} \text{ and } P_1=\frac{24}{30}P_0$$

(ix) The waiting time decreases exponentially as the number of servers increases.



1(b)(i) Using Model 1, the regression equation for Salary (response variable) and Qualification and Gender (predictor variables) is found to be:

$$\text{Salary} = 20009.5 + 0.935253\ Q + 0.224337\ G + \varepsilon$$

Gender is categorical. Keeping the qualification constant (0 for Women and 1 for men), the salary for women and men can be determined as follows:

Salary for women:
E[Salary| Qualification, Gender= 0]= 20009.5 + 0.935253Q + ε

Salary for men:
E[Salary| Qualification, Gender = 1]= (20009.5+0.224337) + 0.935253 Q = 20009.724+ 0.935253 Q + ε

When qualifications are the same, difference in salary for men and women is

20009.724-20009.5 = 0.224.

Thus, on an average, men are paid $224 more than equally qualified women. The regression in Model 1 indicates that men receive slightly more than women. Holding qualification constant, men earn 0.224K, or $224, more than women. An interesting question is whether this difference is statistically significant. Looking at the p-value of the t-statistic, over 0.6329, one would be inclined to say no; the difference is small and could be due to chance.

(ii) Using Model 2, the regression equation for Qualification (response variable) and Salary and Gender (predictor variables) is found to be:

$$\text{Qualification} = -16744.4 + 0.850979\ G + 0.836991\ S + \varepsilon$$

Qualification for women:
E[Qualification | Salary, Gender= 0] = -16744.4 + 0.836991 S + ε

Qualification for men:
E[Qualification | Salary, Gender=1] = -16744.4 + 0.850979 + 0.836991 S + ε
= + 0.836991 S + ε = -16743.54903

When salaries are the same, difference in salary for men and women is

-16744.4 - (-16743.54903) = 0.836991.

Thus, on an average, the index of employee qualification for women is lower than men as much as 0.850979. Holding salary constant, the regression in Model 2 indicates that men are slightly more qualified than women. An interesting question is whether this difference is statistically significant.

Looking at the p-value of the t-statistic, over 0.0532, one would be inclined to say yes; the difference is significant as it is equal to 5.32% and p/2=2.66%. For example, considering confidence level of 95% (i.e. significant level of 5%) this difference is significant.

(iii) We detect inconsistency between two models because Model 1 says that men and women receive more or less the same salary but Model 2 says qualification of men is significantly higher than women.

(iv) Each of these two models have advantages and disadvantages. Model 1 looks practically more rational because salary should be a dependent variable identified by qualification. But t-statistic and p-value of Model 1 are not as good as Model 2. Perhaps the best strategy is that we drop the sex variable from Model 1 and fit another regression model to that. This issue with model one can be due to multicollinearity, too. A correlation test may help with this. As another suggestion, there may be some other variables which are not reflected in the model. Finally, we suggest a model for men and another model for men is designed in which salary is a function of qualification rather than consider sex as a binary (0-1) variable.

QUESTION 2.

2(a)(i)

A. The items may be taken to represent the stages, i.e., A is stage 1, B is stage 2, etc.; there are $N = 4$ stages. State s_n = Unused capacity remaining in the knapsack at the start of stage n

B. Decision $x_n = \begin{cases} 1 \text{ if item } n \text{ is included in the knapsack} \\ 0 \text{ otherwise} \end{cases}$.

Constraint on x_n as a function of s_n: $x_n w_n \le s_n$ (item n cannot be included if there isn't enough space available in the knapsack).

C. State transformation equation $s_{n+1} = g(s_n, x_n) = s_n - x_n w_n$.

Value of the initial state (s_1) = 80 = capacity of empty knapsack.

D. Objective function over all n stages: $\sum_n c(s_n, x_n) = \sum_n x_n v_n$.

Optimal recursion relationship:

$$\begin{aligned} f_n^*(s_n) &= \text{ Optimal cost to go from stage } n \text{ and state } s_n \text{ to stage } N \\ &= \max_{x_n w_n \le s_n} \left\{ x_n v_n + f_{n+1}^*(s_n - x_n w_n) \right\} \end{aligned}$$

2(a)(ii)

***N* = 4:**

s_4	x_4	$f_4(s_4, x_4)$	$f_4^*(s_4)$
< 25	0	0	0
≥ 25	0	0	
	1	50*	50

***N* = 3:**

s_3	x_3	s_4	$c_3(s_3, x_3)$	$f_4^*(s_4)$	$f_3(s_3, x_3)$	$f_3^*(s_3)$
80	0	80	0	50	0 + 50 = 50	
(0,0,_,_)						
	1	80 – 35 = 45	40	50	40 + 50 = 90*	90
50	0	50	0	50	0 + 50 = 50*	
(0,1,_,_)						
	1	50 – 35 = 15	40	0	40 + 0 = 40	50
40	0	40	0	50	0 + 50 = 50*	
(1,0,_,_)						
	1	40 – 35 = 5	40	0	40 + 0 = 40	50
10	0	10	0	0	0	0
(1,1,_,_)						

***N* = 2:**

s_2	x_2	s_3	$c_2(s_2, x_2)$	$f_3^*(s_3)$	$F_2(s_2, x_2)$	$f_2^*(s_2)$
80	0	80	0	90	0 + 90 = 90	
(0,_,_,_)						
	1	80 – 30 = 50	75	50	75 + 50 = 125*	125
40	0	40	0	50	0 + 50 = 50	
(1,_,_,_)						
	1	40 – 30 = 10	75	0	75 + 0 = 75*	75

***N* = 1:**

s_1	x_1	s_2	$c_1(s_1, x_1)$	$f_2^*(s_2)$	$f_1(s_1, x_1)$	$f_1^*(s_1)$
80	0	80	0	125	0 + 125 = 125	
	1	80–40 = 40	70	75	70 + 75 = 145*	145

The optimal solution is $(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$, i.e., put A and B in the knapsack. The total value of the items is 145.

2(b)(i) We often have relationships between dependent and independent variables that are not linear. The nonlinear relationship between dependant and independent variables can be detected by looking at the scatter plot of two variables. Therefore, it is suggested that before performing any regression analysis it is important to eyeball the relationship (whether it is linear or not) by checking the scatter plot. In summary the scatter plot shows that it seems reasonable to assume that the relationship is linear.

(ii) The r^2 statistic (the coefficient of determination) indicates how well an estimated regression function fits the data. It measures the proportion of the total variation in the dependent variable y around its mean that is accounted for by the independent variable in the estimated regression equation. However, the **r^2 statistic** is a biased estimate based on your sample; it tends to be too high. This bias is a reason why adjusted R-squared should be used. Another potential problem with a high **r^2 statistic** is the possibility of overfitting. An overfit model has too many independent variables. The regression model may be tailored to fit the particular data set and may not fit a different data from the same population. You may want to check the p-values for independent variables.

(iii) There are different measures of risk one can use for portfolio management. Some of them are as follows:

- **Variance** describes "risk of not being average".
- **Semi-variance** describes "risk of loss". It equals ½ of variance if symmetric distribution.
- **Regret** is "risk of being different from the observed outcome".
- A very risk-averse risk measure is to consider the **worst possible outcome**, discarding its probability.
- **5th percentile** describes "the best outcome amongst the worst 5% of outcomes". It is related to value at risk (VaR).
- **Average outcome** amongst all below 5th percentile. It is also related to conditional value at risk CVaR.

(iv) Assume that there are two assets x and y with mean returns r_x and r_y, variances Var(x) and Var(y), and covariance Cov(x; y). Then, one version of the mean-variance model can be defined as follows: Given a return level r, find a portfolio v generated from assets x and y to minimize the risk Var(v) such that $r_v \geq r$:

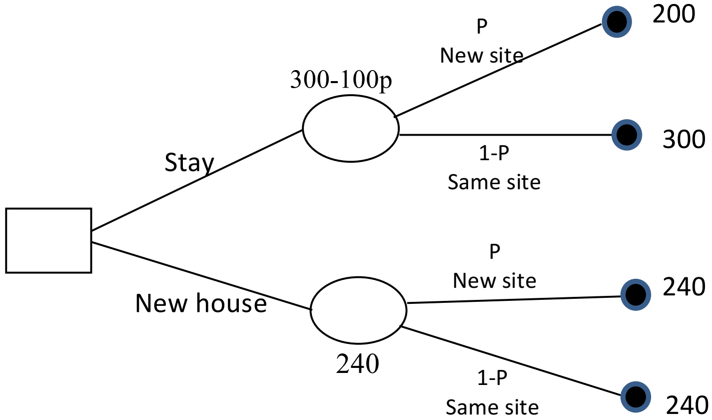
$$\begin{aligned} \min_{1 \geq \alpha \geq 0} \quad & \alpha^2 \mathrm{Var}(x) + (1-\alpha)^2 \mathrm{Var}(y) + 2\alpha(1-\alpha)\mathrm{Cov}(x,y) \\ \text{s.t.} \quad & \alpha r_x + (1-\alpha) r_y \geq r. \end{aligned}$$

This is a simple optimization problem, which can be solved with many optimization software packages, for example, Solver, which is an Excel Add-in. Another version of the mean-variance model can be defined as follows: Given a risk level s^2, find a portfolio v generated from assets x and y to maximize the return r_v such that Var(v)$\leq s^2$:

$$\begin{aligned} \max_{1 \geq \alpha \geq 0} \quad & \alpha r_x + (1-\alpha) r_y \\ \text{s.t.} \quad & \alpha^2 \mathrm{Var}(x) + (1-\alpha)^2 \mathrm{Var}(y) + 2\alpha(1-\alpha)\mathrm{Cov}(x,y) \leq s^2. \end{aligned}$$

QUESTION 3

(a)(i)



(ii)
EMV= Min{300-100p, 240}. We use minimisation because the payoffs are cost.

300-100p=240 Then, p=0.6.

In other words, for p<60% moving to the new house and for p>60% staying are the best decisions respectively. For p=60% both decisions are equal with the same expected cost.

(iii) We need to calculate $p(s_1|a_1)$, $p(s_2|a_1)$, $p(s_1|a_2)$ and $p(s_2|a_1)$ using the Bayesian rule.

$$P(s_1|a_1) = \frac{8p}{1+7p}, P(s_2|a_1) = \frac{1-p}{1+7p},$$
$$P(s_1|a_2) = \frac{2p}{9-7p}, P(s_2|a_2) = \frac{9-9p}{9-7p}.$$

Next, we calculate the expected payoff for stay given that the boss is favourable:

$$\frac{300 + 1300p}{1 + 7p}$$

The payoff for moving to the new house is 240; therefore, I would stay when

$$\frac{300+1300p}{1+7p} \leq 240 \Rightarrow 3/19 \leq p.$$

(iv)
Advantages:

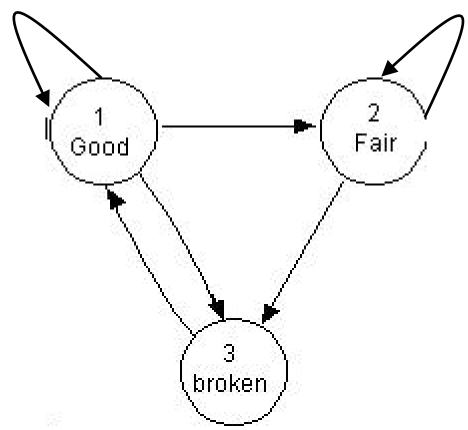
- Simple, visual method to solve the decision-making problem systematically.
- We can consider uncertainties.
- Sensitivity analysis on payoff and probabilities can be applied.

Disadvantages:

- Assumes risk neutrality and ignore risk preferences.
- Difficult to estimate probabilities.

(b)(i) The states are Good (G), Fair (F) and Broken (B). The transition matrix is as follows:

$$P = \begin{array}{c} \begin{array}{ccc} G & F & B \end{array} \\ \begin{bmatrix} 0.85 & 0.1 & 0.05 \\ 0 & 0.7 & 0.3 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$



(ii) State i and j **communicate** if i is accessible from j and j is accessible from i. G and F are accessible from each other. B and F are accessible from each other. G and B are accessible from each other. So, all pairs of states communicate. Communicating states form **classes**. Therefore, there is only one class which is {G, F, B}.

(iii) A process has **periodic behaviour** if the process can only recur to state i after t,2t,3t,… steps. There exists t: if n Not in {t,2t,3t} then $P_{ii}^{(n)} = 0$.

Period of a state i is equal to the greatest common divisor of n such that $P_{ii}^{(n)} > 0$.
A state with period 1 is called **aperiodic.**

- A state with period 1 is called **aperiodic.**
- State i is aperiodic if and only if there exists N such that $P_{ii}^{(n)} > 0$ for all $n > N$.
- State i is aperiodic if and only if there exists N such that $P_{ii}^{(N)} > 0$ and $P_{ii}^{(N+1)} > 0$.

Therefore, all states G, F, and B are in the same class with period of 1. So, the transition matrix is aperiodic.

(iv) An irreducible and aperiodic Markov chain has a steady state distribution.

- A stochastic matrix is said to be **irreducible** if each state is accessible from each other state.

The steady-state equations are as follows:

$$\pi P = \pi$$
$$\sum_{i=1}^{m} \pi_i = 1$$

π_G=0.674 157 303 → 1/ π_G=1.483
π_F=0.224 719 101 →1/π_F=4.44875
π_B=0.101 123 596 →1/π_B=9.8888

1.483+4.44875+9.8888=15.82055 years.

Annual operational cost = ((1.483×£1000)+ (4.44875×£1500) + (9.8888×£0))/ 15.82055
=£421.895

Annual fixed cos = (£6000/1.483) + (£2000/4.44875) + (£0/9.8888)
=£4045.85 + £449.56+0=£4495.41

Overall annual cost = Annual operational cost + Annual operational cost
= £421.895+£4495.41=£4917.309

(v)

Solution 1

$$\begin{array}{cc} & \begin{array}{ccc} G & F & B \end{array} \\ P = & \begin{bmatrix} 0.85 & 0.1 & 0.05 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

π_G=0.869 565 217→ 1/ π_G=1.15
π_F=0.086 956 522→1/π_F=11.4999=~11.5
π_B=0.043 478 261→1/π_B=22.99999=~23

Then we can do the same calculations we did with **(b)(iv).**

Solution 2:

$$\begin{array}{cc} & \begin{array}{cc} G & B/F \end{array} \\ P = & \begin{bmatrix} 0.85 & 0.15 \\ 1 & 0 \end{bmatrix} \end{array}$$

π_G=0.869→ 1/ π_G=1.15
$\pi_{B/F}$=0.131→1/$\pi_{B/F}$=7.634

1.15+7.634=8.784 years.
Annual operational cost= ((1.15×£1000)+ (7.634×£0))/ 8.784=£130.01 (so the operational cost is lower than before)

Annual fixed cost= (£6000/1.15) + (£0/7.634)=£4045.85 + £449.56+0=£5217.39 (obviously the fixed cost is higher)

Overall annual cost = Annual operational cost + Annual operational cost= £130.01+£5217.39=£5347.40

Therefore, the overall cost is £5347.40 - £4917.309=£430.092 higher if we replace the fair car with a good one.