Cribs for the 3E3 Modelling Risk Exam Paper (2013-2014)

Q1(a)(i) The data is $\lambda = 1/10$ and $\mu = 1/6$. For the $M/M/1$ queue,

$$L = \frac{\lambda}{\mu - \lambda}.$$

Note this formula can be derived from the queueing formulas in the data sheet. Therefore, we have

$$L = \frac{3}{2}.$$

Q1(a)(ii) Note that Little's law states that

$$W = \frac{L}{\lambda}, \qquad L_q = \lambda W_q,$$

and also

$$W_q = W - \frac{1}{\mu}.$$

We have

$$L_q = \lambda W_q = \lambda(W - \frac{1}{\mu}) = \lambda(\frac{L}{\lambda} - \frac{1}{\mu}) = \frac{9}{10}.$$

Q1(a)(iii) Little's law states that

$$W = \frac{L}{\lambda}, \qquad L_q = \lambda W_q.$$

On average, the number of customers at the station is equal to the arrival rate multiplied by the average time for a customer spending at the station. On average, the number of customers in the queue is equal to the arrival rate multiplied by the average time for a customer waiting for the pump becoming available.

Q1(b) The service rates for three servers are

$$\mu_1 = \frac{1}{30}, \qquad \mu_2 = \frac{1}{20}, \qquad \mu_3 = \frac{1}{15}.$$

The aggregate service rate is

$$\mu = \mu_1 + \mu_2 + \mu_3 = \frac{9}{60}.$$

It is known that if n independent exponentially distributed servers with service rates $\mu_1, \ldots, \mu_n$ are busy then the time until the next service completion is exponentially distributed with parameter $\mu = \mu_1 + \ldots + \mu_n$. Therefore, the expected time for the next customer to complete the service is $\frac{1}{\mu} = \frac{60}{9}$. Due to the memoryless property, the expected remaining time until the next service completion is still $\frac{1}{\mu} = \frac{60}{9}$.

Q1(c) The mathematical equation for CAPM is

$$r_i - r_f = \beta_i(r_M - r_f)$$

where r_i is the rate of return for asset i, r_f is the rate of return for the risk-free asset, r_M is the rate of return for the market portfolio which is efficient, and β_i is the correlation coefficient between asset i and market portfolio M. The mathematical equation for the capital market line is

$$r_v - r_f = \frac{\sigma_v}{\sigma_M}(r_M - r_f),$$

where v is a portfolio on the capital market line, σ_v and σ_M are standard deviations of portfolio v and market portfolio M, respectively. The capital market line relates the expected rate of return of an efficient portfolio to its standard deviation, but it does not show the expected rate of return of an individual asset relates to its individual risk. CAPM states that the expected excess rate of return of an asset is proportional to the expected excess rate of return of the market portfolio with proportionality factor of beta.

Q1(d) In the least square method, we aim to minimize the total squared errors:

$$\min_{a,b} f(a,b) = \sum_{i=1}^{n}(a + bx_i - y_i)^2.$$

The objective function is convex. Therefore, the optimal value of the least square problem is obtained at its stationary point. Set the first order derivatives of $f(a,b)$ to be equal to zero:

$$\frac{\partial f}{\partial a} = 2\sum_{i=1}^{n}(a + bx_i - y_i) = 0.$$

$$\frac{\partial f}{\partial b} = 2\sum_{i=1}^{n}(a + bx_i - y_i)x_i = 0.$$

The first equation shows that

$$a + b\bar{x} = \bar{y}, \qquad a = \bar{y} - b\bar{x}.$$

Plugging a into the second equation, we obtain

$$b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})x_i}{\sum_{i=1}^{n}(x_i - \bar{x})x_i}.$$

Q1(e) The two histograms A and B represent two different designs for a project. It seems that we can draw the following conclusions from the histograms. There are 7 different outcomes/profit-values for design A and 6 different outcomes for design B. The average profit for design A is higher than design B. The risk (variance) for design A is larger than that for design B. Design A has a larger chance to lose more than 90 units than design B. Design A has a larger chance to make a profit of 490 units or 90 units than design B. Overall, design B is better than design A because it has a larger average and a smaller risk.

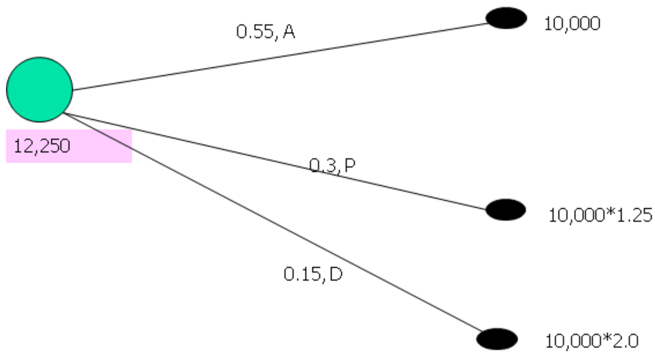

Figure 1: The decision tree when the firm inspects exactly one supplier.

Q2(a)(i) See Figure 1.

The expected cost for the firm is £12,250.00.

Q2(a)(ii) See Figure 2.

The expected cost for the firm is £10,455.625.

Q2(a)(iii) Intuitively, the firm can eventually identify a supplier which can provide Acceptable components if they are allowed to inspect an infinite number of suppliers because 55% of suppliers provide Acceptable components.

Let $p = 0.55$ and C_m be the expected cost for the firm if they can inspect at most m suppliers. Then we have

$$C_{m+1} = p \times 1 \times 10,000 + (1-p) \times C_m.$$

By induction, we have

$$\begin{aligned}
C_{m+1} &= p \times 1 \times 10,000 + (1-p) \times C_m \\
&= p \times 1 \times 10,000 + (1-p) \times [p \times 1 \times 10,000 + (1-p) \times C_{m-1}] \\
&= 10,000p + 10,000p \times (1-p) + (1-p)^2 C_{m-1} \\
&= \cdots \\
&= 10,000p + 10,000p \times (1-p) + \cdots + 10,000p(1-p)^{m-1} + (1-p)^m C_1 \\
&= 10,000p(1 + (1-p) + \cdots + (1-p)^{m-1}) + (1-p)^m C_1 \\
&= 10,000p\tfrac{(1-(1-p)^m)}{1-(1-p)} + (1-p)^m C_1 \\
&= 10,000(1-(1-p)^m) + (1-p)^m C_1.
\end{aligned}$$

Let m tend to infinity, we have

$$\lim_{m \to \infty} C_{m+1} = 10,000.$$

Q2(a)(iv) When the firm can inspect an infinite number of supplier, the firm acquires more and more information about the component quality from all suppliers. Additional information (without any cost in this case) can potentially increase the project value. Indeed, the

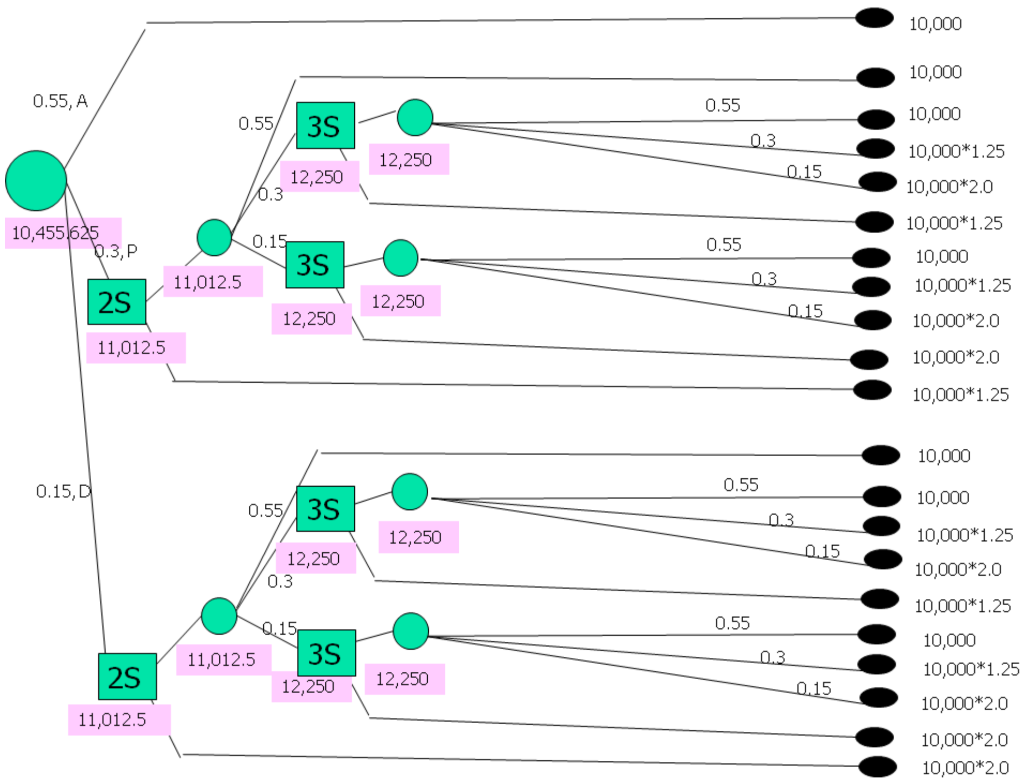

Figure 2: The decision tree when the firm inspects at most three suppliers.

expected value of perfect information is 12,250 - 10,000 = £2,250. Furthermore, the expected value of sample information from inspecting at most three suppliers is 12,250 - 10,455.625 = 1,794.375 £.

Q2(b)(i) In a Markov chain, the period of a state i is equal to the greatest common divisor of n such that $P_{ii}(n) > 0$, where $P_{ii}(n)$ is the n-step transition probability from state i to state i. A state with period 1 is called aperiodic.

Q2(b)(ii) Suppose there exists N such that $P_{ii}(N) > 0$ and $P_{ii}(N+1) > 0$. Then for any $n > N \times (N+1)$, we have $P_{ii}(n) > 0$. Here is a proof. Let i satisfy $0 \le i \le N$ and $n = N \times (N+1) + i$. Then we have

$$n = N \times (N+1) + i = N \times (N+1-i) + (N+1) \times i.$$

Because $P_{ii}(N) > 0$ and $P_{ii}(N+1) > 0$, the above equation shows that

$$P_{ii}(n) > 0.$$

Similarly, we can prove that for any $n > N \times (N+1)$, $P_{ii}(n) > 0$. Therefore, the period for state i is 1.

Q2(b)(iii) For a Markov process $\{X_t\}$, the Markovian property states that

$$P(X_{t+1} = j | X_0 = i_0 \ldots, X_t = i_t) = P(X_{t+1} = j | X_t = i_t).$$

In words: The probabilities that govern a transition from state i at time t to state j at time $t+1$ only depend on the state i at time t and not on the states the process was in before time t.

Q2(c) Sensitivity analysis is a quantitative approach to test how the system output changes in change of an input parameter when all other parameters are fixed.

Sensitivity analysis can inform us whether or not the system output is very sensitive to an input parameter. If yes, we should replace a fixed value for the input parameter by a probability distribution in a Monte Carlo simulation model.

The drawbacks of sensitivity analysis are: It cannot test the joint effects of multiple uncertain inputs; It ignores dependencies between uncertain inputs; and it does not take probabilities into account.

Q2(d)(i) We define the smoothed average at time t to be

$$E_t = \alpha X_t + (1-\alpha)E_{t-1} = E_{t-1} + \alpha(X_t - E_{t-1}).$$

The (one-step) forecast is

$$F_{t+1} = E_t$$

The 2-step forecast is

$$F_{t+2} = F_{t+1} + \alpha(F_{t+1} - E_t) = F_{t+1}$$

and thus $F_{t+k} = F_{t+1}$ for all $k \geq 1$.

Q2(d)(ii) We can iterate the above recurrence relation to get

$$\begin{aligned} E_t &= \alpha X_t + (1-\alpha)(\alpha X_{t-1} + (1-\alpha)E_{t-2}) \\ &= \alpha X_t + (1-\alpha)\alpha X_{t-1} + (1-\alpha)^2 E_{t-2} \\ &= \alpha X_t + (1-\alpha)\alpha X_{t-1} + (1-\alpha)^2 \alpha X_{t-2} + \cdots \\ &= \alpha \sum_{i=0} (1-\alpha)^i X_{t-i}. \end{aligned}$$

The forecast of single exponential smoothing takes all past observations into account by giving more weights to more recent observations (the weight decreases exponentially back in time).

Q3(a)(i) Let (x_1, x_2, x_3, x_4) be the long-run probability distribution for the status of crates (R, G, F, and D). Then it satisfies the following system of equations:

$$(x_1, x_2, x_3, x_4)P = (x_1, x_2, x_3, x_4),$$

and

$$x_1 + x_2 + x_3 + x_4 = 1,$$

where P is the transition probability matrix. More precisely, we have

$$\begin{aligned}
&x_4 = x_1 \\
&0.8x_1 + 0.6x_2 = x_2 \\
&0.2x_1 + 0.4x_2 + 0.5x_3 = x_3 \\
&0.5x_3 = x_4.
\end{aligned}$$

Solving the system of equations, we have

$$x_1 = \frac{1}{6}, x_2 = \frac{1}{3}, x_3 = \frac{1}{3}, x_4 = \frac{1}{6}.$$

Q3(a)(ii) The long-run probability distribution for the system indicates that with a chance of 1/6, a crate needs to be rebuilt, with a chance of 1/3, a crate is in good condition, with a chance of 1/3, a crate is in fair condition, and with a chance of 1/6, a crate is damaged. Alternatively, it also means that at any point in time, 1/6 of all crates are being rebuilt, 1/3 of all crates are in good condition, 1/3 of all crates are in fair condition, and 1/6 of all crates are damaged. If the system is in a steady state in this week, then it will be in a steady state in the next week. That is the following

$$(x_1, x_2, x_3, x_4)P = (x_1, x_2, x_3, x_4)$$

holds.

Q3(a)(iii) The two-step transition probability matrix is

$$P \times P = \begin{pmatrix} 0 & 0.48 & 0.42 & 0.1 \\ 0 & 0.36 & 0.44 & 0.2 \\ 0.5 & 0 & 0.25 & 0.25 \\ 0 & 0.8 & 0.2 & 0 \end{pmatrix}.$$

Therefore, if a crate is in good condition this week, with a chance of 0.36, it will be in good condition in two weeks.

Q3(a)(iv) The expected weekly cost of both rebuilding and loss of production efficiency is

$$2.5 \times \frac{1}{6} + 1.85 \times \frac{1}{6} = \$\frac{4.35}{6} = 0.725,$$

per crate.

Q3(a)(v) In the new plan, there are only three states: R, G, and F and the transition probability matrix is

$$P = \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0 & 0.6 & 0.4 \\ 1 & 0 & 0 \end{pmatrix}.$$

The new steady-state probability distribution is $(x_1, x_2, x_3) = (1/4, 1/2, 1/4)$. The new weekly cost is

$$2.5 \times \frac{1}{4} = \$\frac{2.5}{4} = 0.625,$$

per crate.

Q3(b)(i) When we estimate the population proportion, the meaning of the 95% confidence interval is the following. 95% of all 95% confidence intervals for the population proportion contain the population proportion. In another word, statistically, 5% of the 95% confidence intervals do not contain the population proportion.

Q3(b)(ii) The mathematical formulas for 95% confidence intervals are

$$[q - z\text{STEP}, q + z\text{STEP}],$$

where q is the sample proportion, $z = 1.96$ is the critical value for the confidence level of 95%, and $\text{STEP} = \sqrt{p(1-p)/n}$, p is the population proportion, and n is the sample size. The width of the 95% confidence interval is $2z\text{STEP}$. Clearly, the larger the sample size is, the smaller the width of the 95% confidence interval. It further shows that a large sample size gives more accurate estimation of the population proportion because the width of 95% confidence interval is decreasing in sample size n.

Q3(c)(i) The regression equation is

$$y = -49.531646627214 - 24.7621701355745x_1 + 907.269903225928x_2,$$

where x_1 represents ERA as the first independent variable, x_2 represents AVG as the second independent variable, and y represents the number of wins as the dependent variable, - 24.7621701355745 is the slope for ERA, which represents the change for the number of wins for a unit increase of ERA given that AVG holds unchanged, and a similar meaning can be given for AVG.

Q3(c)(ii) When $x_1 = 4.0$, $x_2 = 0.26$,

$$y = -49.531646627214 - 24.7621701355745x_1 + 907.269903225928x_2 = 87.31.$$

A 95% prediction interval is

$$[y - 1.96S_e, y + 1.96S_e] = [76.99, 97.63],$$

where $S_e = 5.265$ is the standard error of the regression model. At the 95% confidence level, the number of wins for a team with an ERA of 4.0 and an AVG of 0.26 is between 77 and 98 approximately.

Q3(c)(iii) If the confidence level is 80%, the critical value is $z = 1.28$ based on the standard normal distribution table. An 80% confidence interval for the slope of ERA is $[b - zs_b, b + zs_b] = [-30.1688, -19.3556]$ approximately, where $b = -24.76217$ and $s_b = 4.22391256$. At the 80% confidence level, the decrease for the number of wins for a team is between 19 and 30 for a unit increase of ERA.

Note that some marks are given even if a wrong critical value is used, which leads to an incorrect 80% confidence interval.

Q3(c)(iv) We look at the following attributes in Summary Outputs when we compare the simple and multiple regression models.

* A larger R-square statistic implies that the regression model fits the data better. The multiple regression model is better than the simple regression model.
* A smaller standard error gives a higher prediction power. The multiple regression model is better than the simple regression model.
* In order for an independent variable to be significant in the regression model, a larger value of the t-statistic is preferred. At the 95% confidence level, all independent variables are significant in both multiple and simple regression models. Note that it is fine to investigate the p-value or confidence intervals.