EGT2
ENGINEERING TRIPOS PART IIA

Module 3E3

MODELLING RISK – CRIB

1. (a)

(i) Arrival: Customer coming to order a drink. Departure: Drink served to the customer. Queue: Customer waiting for others to be served.

(ii) Arrival rate = 0.167 customers/minute; Service time = 3 minutes. Then, service rate, $\mu = 1/3 = 0.333$ customers per minute. To find the minimum number of baristas needed, we can analyse the queuing system using Little's Law: $L = \lambda W$, where:

- L is the average number of customers in the system (including those being served)
- λ is the average arrival rate of customers
- W is the average waiting time per customer in the system

We similarly define L_q and W_q for the above parameters related to the queue rather than the system.

Case of s = 1
$\mu = 0.333$, utilisation is $\lambda/(\mu s) = 0.167/0.333 = 0.50$
Looking up in annexed table we have that $L_q = 0.50$
then $W_q = L_q / \lambda$; $W_q = 0.5/0.167 = 3$ minutes

Case of s = 2
$\mu = 0.333$, utilisation is $\lambda/(\mu s) = 0.167/0.667 = 0.25$
Looking up in annexed table with L_q in [0.0167, 0.0593]
then $W_q = L_q / \lambda < 0.0593/0.167 = 0.35$ minutes = 21 seconds

Having 2 baristas, the waiting time in queue is less than 21 seconds.

(iii) We have 2 independent systems, regular system with 2 baristas (3 minutes service time each) and priority system with 1 barista (2 minutes service time).

- For the regular system if all the parameters remain (2 baristas) the average waiting time on queue is less than 21 seconds. Or in case of having 1 barista W_q is 3 min as computed at (i).
- For the priority system we have that $\lambda = 0.167$, $\mu = 0.5$ and utilisation is 0.167/0.5 = 0.334 then looking up in annexed table L_q in [0.1286, 0.1885] then W_q is in the interval of 46 s and 1.12 min.

Having 2 independent systems, regular customers wait in queue less than priority customers in case of maintaining 2 baristas for the regular system. Otherwise, having 1 barista at each queue, the priority improves the waiting times for those customers.

(b)

(i) To calculate the 3-month simple moving average (MA), we'll take the average of sales for every consecutive group of 3 months. Then, forecasting the period t will depend on observations in t-1, t-2, and t-3. The solutions can only be computed from April onward.

- MA sales in April = (50 + 60 + 55) / 3 = 165 / 3 = 55
- MA sales in May = (60 + 55 + 70) / 3 = 185 / 3 = 61.67
- MA sales in June = (55 + 70 + 65) / 3 = 190 / 3 = 63.33
- MA sales in July = (70 + 65 + 80) / 3 = 215 / 3 = 71.67

(ii) Exponential smoothing forecasts are calculated using the formula:
$F_t = \alpha Y_{t-1} + (1- \alpha) F_{t-1}$
Considering the forecast for February the observed in January, F2 = 50. The following forecasting periods are computed using the above formula:
F3 = 0.3 * 60 + (1 - 0.3) * 50 = 53
F4 = 0.3∗55 + 0.7∗53 = 53.6
F5 = 0.3∗70 + 0.7∗53.6 = 58.52
F6 = 0.3∗65 + 0.7∗58.52 = 60.464
F7 = 0.3∗80 + 0.7∗60.464 = 66.3248

The meaning of the smoothing factor is weight of the observations depending on the distance in time with the forecast. Small values of the smoothing factor will make predictions depending less on the last changes in the time series data. That is, less responsive to the latest observations and smoother in the prediction.

(iii) Time series analysis involves studying and modelling the patterns, trends, and behaviours of data points collected over time. It is widely used in various fields, such as finance, economics, weather forecasting, and stock market analysis.
Key Terms:

- Time Series: A time-ordered sequence of data points or observations collected at regular intervals over time.
- Trend Components: Trends represent long-term movements or patterns in a time series. They can be upward (increasing), downward (decreasing), or flat. Trends provide insights into the underlying structure of the data.
- Cyclical Patterns: Cyclical patterns are repetitive and predictable movements in a time series that are not strictly tied to seasonality. These cycles are often of longer duration and may not have fixed periodicity. Economic cycles, for example, exhibit cyclical patterns.

(c)

A Markov chain is a stochastic process that undergoes transitions between a sequence of states according to certain probabilistic rules. The key feature of a Markov chain states that the probability of transitioning to any state depends solely on the current state and time elapsed, regardless of how the system arrived at its current state. In other words, the future behaviour of the process is independent of its past, given its present state. This memoryless property is often referred to as the "Markov property".

Example of a Markov Chain is a simple weather model with states representing weather conditions such as "sunny," "cloudy," and "rainy." The transition probabilities might capture the likelihood of weather changes from day to day based on historical data.

2. (a)

(i) An example of risk-return chart is the following shown at Figure 1

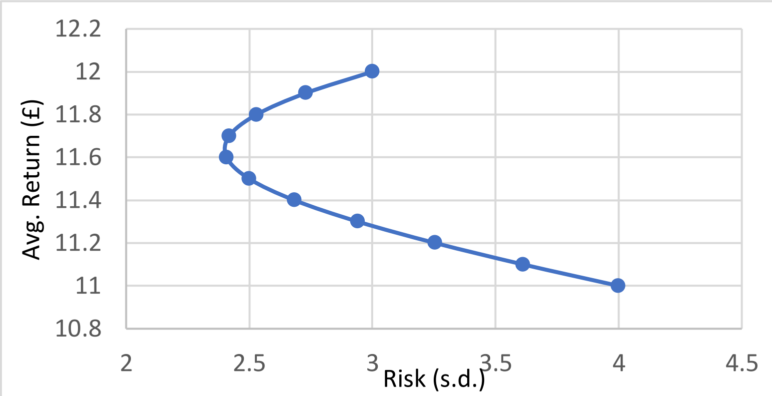

Fig. 1

Investment options can be classified as follows based on risk:

- Low risk, low return: Typically includes safer investments like government bonds or savings accounts.
- Moderate risk, moderate return: Balanced funds or diversified portfolios fall into this category.
- High risk, high return: Stocks, venture capital, or commodities often fall into this category.

[Example of answer] As for my preferred investment strategy, it depends on various factors such as financial goals, time horizon, and risk tolerance. Generally, I lean towards a diversified portfolio with a mix of assets to spread risk while aiming for reasonable returns over the long term. This approach allows for potential growth while mitigating the impact of market fluctuations.

(ii) $X \sim N(12, 9)$, $Y \sim N(11, 16)$.

$V = X/2 + Y/2$, as independents we have that
$E[V] = E[X/2] + E[Y/2] = E[X]/2 + E[Y]/2 = 11.5$

Similarly,
$Cov(X,Y) = 0$
$S^2_V = S^2_X/4 + S^2_Y/4 = 25/4$, and the standard deviation is $S_V = 2.5$

Answering the question about the probability of a 50-50 portfolio losing 25% or more of its value, we just use the new variable V. $V \sim N(11.5, 2.5^2)$,

The 25% of the initial value (£10/share) is £2.50, bringing value of v for the question of £7.50. Then,

$P(V < 7.50) = P(Z < (7.50-11.5)/2.5) = P(Z < -1.6) = 0.0548$ (from the table of the standard normal distribution as $Z \sim N(0, 1)$)

Compare the result above to a 100% investment in Y, we have that:

$P(Y < 7.50) = P(Z < (7.50-11)/4) = P(Z < -0.875) = 0.18$ (approx., from the table of the standard normal distribution as $Z \sim N(0, 1)$)

The probability $P(Z < -1.6) = 0.0548$ is lower than $P(Z < -0.875) = 0.18$. Hence, the probability of losing more than 25% of the 50-50 portfolio's value is lower than in the case of a 100% investment in Y. The 50-50 portfolio is less risky due to diversification.

(b)

(i) Let's work with the decision tree of Figure 2.

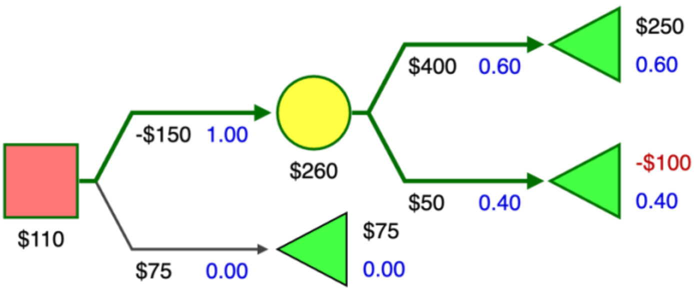

Fig. 2

Financial feasibility:

Expected Value "launching the product": $(0.6 \times 400{,}000) + (0.4 \times 50{,}000) - 150{,}000 =$ £110,000
Expected Value "no launching the product": £75,000.

Financially, invest on going forward and launching the product has a higher expected value.

(ii) Probability threshold:

For "no launching the product" to be optimal, its expected value should be greater than that of "launching". Given P(positive market response) = p
$400p + 50(1-p) - 150 \leq 75$ (numbers in thousands)
$p \leq 0.5$

Therefore, at a positive market response probability below 0.5, not launching the prototype to the market would be financially wiser.

(iii) Cost threshold:

For "launching" to be optimal decision, its expected value should remain greater than the value for no launching the product, that is:
$0.6 \times 400 + 0.4 \times 50 - (150 + X) \geq 75$ (numbers in thousands)

110,000 – X ≥ 75; X ≤ 35,000. The additional cost should not exceed £35,000 for the "launching" option to be financially viable.

(iv) Given the following variables and states from the question:
A = evaluation results with values: a1 = desirable, a2 = undesirable
S = market status with values: s1 = favourable, s2 = unfavourable conditions.

The question states that:
P(A=a1 | S = s1) = 0.8; P(A = a2| S = s2) = 0.9
P(A = a2| S = s1) = 0.2; P(A = a1 | S= s2) = 0.1

We also know that P(S = s1) = 0.6 and P(S = s2) = 0.4

Using Bayes' theorem:

$$\mathrm{P(S = s1| A = a1)} = \frac{P(A = a1 \mid S = s1)P(S = s1)}{P(A = a1 \mid S = s1)P(S = s1) + P(A = a1 \mid S = s2)P(S = s2)}$$

P(S = s1| A = a1) = 0.8×0.6/(0.8×0.6 + 0.1×0.4) = 0.92

The value of the information is computed then as the difference:
Expected value with info – Expected value without info (from the result in (i))
(0.92×400,000 + 0.08×50 – 150,000) – 110,000 = 368,040 – 110,000 = 108,040

As the maximum that we will be willing to pay for the information is £108,000

(c)

The principal pitfalls in data analytics are:

- Bias in data selection: Sample may not be representative of the population. What was a good predictor for the selected sample may not be a good predictor for the population.
- Bias in data collection: If the data used to build a model involves decisions made by biased individuals, then those biases can be `baked' in.
- Outliers can also pose an issue as they have a disproportional impact on the way the algorithm will make predictions.

Other examples of major pitfalls are the existence of cofounding variables, misinterpretation of correlation/causation, and ethical concerns.

3. (a)

(i) Table 1 shows the spaces with the required values:

Source	Sum of Squares	Deg. of freedom	Mean Square	F-Statistic	p-value
Regression	120,000	1	120,000	1.5	P < 0.05
Residual	80,000	18	4,444.44		
Total	200,000	19			

Table 1

Null Hypothesis for the F-test:

The null hypothesis for the F-test in this context is that the regression model does not explain a statistically significant portion of the variance in the sales variable. In other words, the null hypothesis states that the slope of the regression line is equal to zero and there is no linear relationship between advertising expenditure and sales.

Sample Size:

Given the degrees of freedom for regression (1) and residual (18), the total degrees of freedom for the analysis are the sum (1 + 18) = 19. Based on the formula for sample size in regression analysis (n = total degrees of freedom + 2), the sample size used in this study is n = 19 + 2 = 21.

(ii) R-squared and Interpretation:

The coefficient of determination (R-squared) can be calculated by dividing the sum of squares due to regression by the total sum of squares: 120,000/200,000 = 0.6. This indicates that 60% of the variance in sales is explained by the regression model. In the context of the study, it suggests that advertising expenditure has a positive correlation to sales.

Hypothesis Test for Regression Coefficient:

The null hypothesis for this test is that the regression coefficient (slope) is equal to zero, suggesting a non-significant linear relationship. The alternative hypothesis is that the coefficient is not equal to zero, significant linear relationship. A low p-value (typically less than 0.05, as it is the case of this question) would lead to rejecting the null hypothesis and concluding that the regression coefficient is significantly different from zero, supporting the existence of a linear relationship between advertising expenditure and sales.

(iii) Confidence Interval for the Slope:

$$\beta' \pm t\_(\alpha/2, n\text{-}2) * SE(\beta')$$

$SE(\beta') = \sqrt{(\text{Mean Square for Residual} / (n-2))} = \sqrt{(4{,}444.44 / (21-2))} \approx 4.71$

For a 95% confidence interval, $\alpha = 0.05$, so $\alpha/2 = 0.025$. With 17 degrees of freedom (n-2 = 21-2), the critical t-value from a t-distribution table is approximately 2.110 (see t-student distribution table)

(b)

(i) The salvage value (v) represents the revenue Taylor can get from selling excess inventory after the initial demand is satisfied. From the newsvendor problem we have that

Underage cost (cu) = price – cost = 15 – 10 = 5
Overage cost (co) = cost – salvage value = 10 – v

Since Taylor targets Q = 55, we have that: F(Q = 55) = cu/(cu + co) = 5/(5 + (10 - v)). We also know that D ~ N(50, 6)

Then, $F(Q = 55) = P(D \leq 55) = P(Z \leq (55 - 50)/6) = P(Z \leq 0.83) = 1 - 0.2033 = 0.7967$ from the table of the standard normal distribution as Z ~ N(0, 1)

By letting $5/(15 - v) = 0.7967$, then $5 = 0.7967 \times 15 - 0.7967v$
$5 = 11.95 - 0.7967v$; v = £8.72

(ii) If the salvage value is set such that the underage and overage costs are equal, it implies that the optimal order is the mean demand forecast. Then Q* = 50. That is, the cost of not meeting demand is the same as the cost of holding excess inventory. In such a scenario, the optimal order quantity would indeed be equal to the expected demand because it minimizes the total cost.

(iii) Now we have: co = 10 – 7 = 3, while cu remains equal than in (i), cu = 5
Critical ratio = cu / (cu + co) = 5 / (5 + 3) = 0.625; thus, from the table below the optimal quantity order is Q* = 80

Q	50	60	70	80	90	100
P(X = Q)	0.15	0.20	0.25	0.18	0.12	0.10
P(X <= Q)	0.15	0.35	0.60	0.78	0.90	1

(c)

Bottleneck is a point in a production process with limited capacity, restricting overall flow and efficiency. Examples of its negative impacts are delays in order fulfilment, decreased customer satisfaction, lost revenue, and higher operational costs. Bottleneck is identified as the step of the process with lowest capacity. Additional resources should be allocated at the bottleneck step and work/tasks should be re-allocated from the bottleneck step to other steps of the production process.

END OF PAPER

THIS PAGE IS BLANK