Version PM/2

EGT2

ENGINEERING TRIPOS PART IIA

Module 3E3

MODELLING RISK - CRIB

1 (a) A job shop consists of three machines and two repair workers. The amount of time a machine works before breaking down is exponentially distributed with a mean of 12 time units. The amount of time it takes a single repair worker to fix a machine is exponentially distributed with a mean of 10 time units.

(i) What is the average number of machines not in use? [25%]

(ii) What proportion of time are both repair workers busy? [25%]

ANSWER: This is a birth-death process with the state space is $\{0,1,2,3\}$, where the states represent the number of machines down. The transition rates are as follows:

$$\lambda_0 = \frac{3}{12},\ \lambda_1 = \frac{2}{12},\ \lambda_2 = \frac{1}{12}, \mu_1 = \frac{1}{10},\ \mu_2 = \frac{2}{10},\ \mu_3 = \frac{2}{10}.$$

Let P_i be the stationary probability for state i, $i = 0,1,2,3$. It follows from the balance equations that

$$P_1 = \frac{5}{2}P_0,\ P_2 = \frac{25}{12}P_0,\ P_3 = \frac{125}{144}P_0.$$

As $P_0+P_1+P_2+P_3 = 1$, we obtain $P_0 = \frac{144}{929} = 0.155$, $P_1 = 0.388$, $P_2 = 0.323$, $P_1 = 0.134$. The average number of machines not in use is $P_1 + 2P_2 + 3P_3 = 1.437$. The proportion of time both repair workers are busy is $P_2 + P_3 = 0.457$.

(b) A racing bike manufacturer (XMB) is using a special seat in its brand-new bike model. The bike sales for this model show a fairly steady demand of 5,600 bikes per year. Traditionally, XMB purchases these seats from a producer in Germany at a price of £8/unit. It costs XMB £100 to place an order. Inventory holding costs are based on an annual interest rate of 20%. Suppose that the seat supplier is offering a quantity discount applied to all units as follows:

Price per seat	**Order quantity, Q**
£8	$Q \leq 800$
£7	$800 < Q < 1000$
£6	$Q \geq 1000$

(i) What is the optimal order quantity in this case? [30%]

(ii) Using your optimal order quantity, what is the total cost? [10%]

ANSWER: First, we calculate our optimal Economic Order Quantities:

$$
\begin{aligned}
Q_8^* &= \sqrt{\frac{2*100*5,600}{0.2*8}} = 837 \\
Q_7^* &= \sqrt{\frac{2*100*5,600}{0.2*7}} = 894 \\
Q_6^* &= \sqrt{\frac{2*100*5,600}{0.2*6}} = 966
\end{aligned}
$$

For an £8 unit cost, $Q_8^* = 837$ is infeasible, so we use $Q^* = 800$ to calculate our Total Cost. For a £7 unit cost, our calculated Q_7^* is feasible (within our bounds). For a £6 unit cost, $Q_6^* = 966$ is infeasible, so we use $Q^* = 1000$.

$$
\begin{aligned}
\text{Total Cost} &= \frac{C_O \times \mathrm{D}}{Q^*} + \frac{Q^*}{2} \times i \times C_u + C_u \times D \\
TC(Q^* = 800, £8) &= \frac{100*5,600}{800} + \frac{800}{2} * 0.2*8 + 8*5,600 = £46,140 \\
TC(Q^* = 894, £7) &= \frac{100*5,600}{894} + \frac{894}{2} * 0.2*7 + 7*5,600 = £40,452 \\
TC(Q^* = 1,000, £6) &= \frac{100*5,600}{1,000} + \frac{1,000}{2} * 0.2*6 + 6*5,600 = £34,760
\end{aligned}
$$

The optimal order quantity is $Q = 1,000$ (at £6 per unit), for a Total Cost = £34,760.

(c) In regression analysis, briefly explain what R^2 is and how it can be used. What is one shortcoming of the R^2 statistic? [10%]

ANSWER: The R^2 statistic (the coefficient of determination), $0 \leq R^2 \leq 1$, 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.

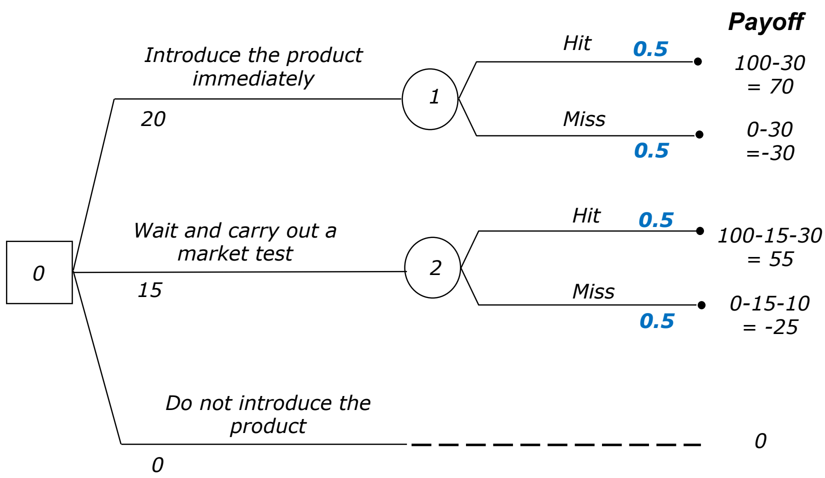
However, the R^2 statistic is a biased estimate based on your sample; it tends to be too high. This bias is a reason why adjusted R-squared should be used. Another potential problem with a high R^2 statistic is the possibility of overfitting. An overfit model has too many independent variables. The regression model may be tailored to fit the particular data set and may not fit different data from the same population. You may want to check the p-values for independent variables.

2 (a) A firm is considering introducing a new product. There are initially three possibilities: (i) Introduce the product immediately; (ii) Wait and carry out a test market; (iii) Do not introduce the product. The firm is undecided about which choice to make because it does not know whether the new product will be a 'hit' (H) or 'miss' (M); introducing a 'hit' product will yield £100K, a 'miss' product nothing.

Product introduction costs £30K. Carrying out a test market costs £15K. A test market will tell for certain whether the product is H or M. If it shows H, the firm will proceed with product introduction; if M they will not introduce the product, which at that stage will cost £10K due to reputational issues. The chance of H is now assessed at 0.5.

(i) Build a decision tree for the firm's decision. [15%]

ANSWER:

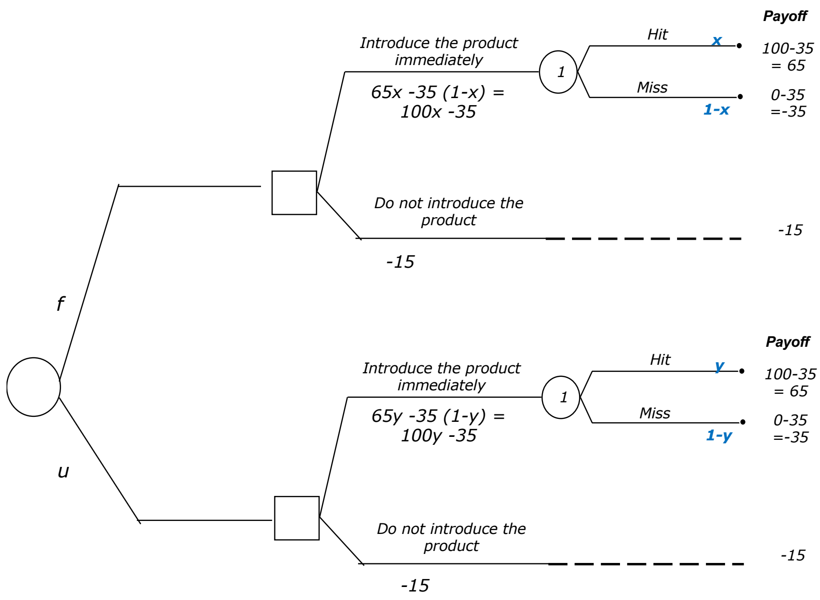


(ii) Based on this decision tree, solve for the firm's optimal decision if the firm wishes to maximize its payoff. [10%]

ANSWER: The firm would introduce the product immediately, with an expected payoff of 20.

(iii) Instead of carrying out a test market themselves, the firm can employ a market research organisation to carry out a test market. The market research organisation can produce a favourable (f) or unfavourable (u) report. These are not infallible, the firm assesses $p(f|H)$ as 0.9 but $p(u|M)$ as 0.7. The organisation charges £5K for its survey. If the firm employs the market research organisation, solve for the firm's optimal decision. Should the firm work with the market research organisation? [25%]

ANSWER:



Using Bayes' Rule:

$$x = P(H|f) = \frac{P(f|H) * P(H)}{P(f|H) * P(H) + P(f|M) * P(M)} = \frac{0.9 * 0.5}{0.9 * 0.5 + 0.3 * 0.5} = \frac{2}{3}$$

$$1 - y = P(M|u) = \frac{P(u|M) * P(M)}{P(u|M) * P(M) + P(u|H) * P(H)} = \frac{0.7 * 0.5}{0.7 * 0.5 + 0.1 * 0.5} = \frac{7}{8}$$

Therefore, the firm should intoduce the product if the market reserach organisation is favourable (with an expected payoff of 95/3), and they should not introduce the product if the market reserach organisation is unfavourable (with an expected payoff of -15).
Since $P(f) = 0.6$ and $P(u) = 0.4$, the expected value of sample information is 95/3 * 0.6 - 15 * 0.4 - 20 = -7, so **NO** the firm should not work with the market research organization.

(b) Briefly explain what Markowitz portfolio analysis is and how it works. [20%]

ANSWER: Markowitz portfolio analysis suggests that it is possible to construct an efficient frontier of optimal portfolios, offering the maximum possible expected return for a given level of risk.

Portfolio weights - the proportions that you invest in each opportunity - allow you to trade off risk against expected return. You can minimise risk subject to a given level of return, or maximise return subject to a given level of risk.

The risk-return profile of portfolios is bounded by the efficient frontier; portfolios that are not on the efficient frontier can be replaced by portfolios that have higher return and lower risk.

(c) Briefly explain what a Markov chain is. How can we characterise a Markov chain? [15%]

ANSWER: A Markov chain is a stochastic process with a finite number, say n, possible states that has the Markov property:

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

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.

A Markov chain is completely characterised by transition probabilities P_{ij} from state i to state j that are stored in an $n \times n$ transition matrix P. Rows of transition matrix sum up to 1. Such a matrix is called a stochastic matrix. Initial distribution of states is given by an initial probability vector.

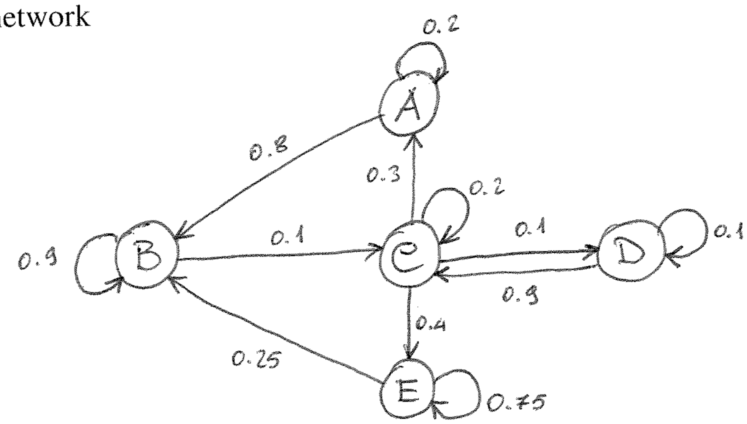
(d) Consider the following transition matrix (numbers in %):

From/To	A	B	C	D	E
A	20	80	0	0	0
B	0	90	10	0	0
C	30	0	20	10	40
D	0	0	10	90	0
E	0	25	0	0	75

Draw the transition network and determine the classes. Does the Markov chain have absorbing classes? [15%]

ANSWER:

Transition network



The Markov chain has only one (absorbing) class.

3 (a) One of your employees has been working on a model to schedule the production of an important part in a least-cost manner. Unfortunately, you spilled coffee on the sheet of paper which contained the solution to the dynamic programming problem (see the table below). You decide to try and reconstruct the solution yourself.

Period t

	1	2	3	4
s_t	200	200	200	200
d_t		29	30	
$Y_t(1)$		258	378	
$Y_t(2)$				
$Y_t(3)$				
$Y_t(4)$				
z_t				
δ				

$h_t =$ for all t

To assist yourself, you jot down the following notes:

- Let n^* equal the number of months worth of demand that production in the first month will cover. For example, if $n^* = 3$, then production in month 1 will meet demand for months 1, 2, and 3.
- Let s_t represent the setup cost.
- Let d_t represent the demand for month t.
- Define z_t as the cost of an optimal schedule for the first t periods.
- Define $Y_t(m)$ as the cost of meeting demand for periods 1 through t by producing in period m $(m \leq t)$ for periods m through t and following an optimal production schedule for the first $(m-1)$ periods.
- Let $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$, where $\delta_t = 0$ if no production in that period, and 1 otherwise.
- No shortages are allowed and there is no production lead time. The holding cost is incurred at the end of each period and is constant over all periods.

 (cont.

(i) Using the notation on the previous page and defining $C_t(m)$ as the cost of meeting demand for period m through t by producing in period m ($m \leq t$), write closed-form expressions for $Y_t(m)$ and z_t. [10%]

ANSWER:

$$Y_t(m) = C_t(m) + Z_{m-1},$$

where $C_t(m)$ is the cost of meeting demand for periods m through t by producing in period m ($m \leq t$), and

$$z_t = \min_{m \leq t} Y_t(m).$$

(ii) What is the value of h_t? [10%]

ANSWER: $Y_3(1) - Y_2(1) = 378 - 258 = 120 = 2 * h * d_3 \Rightarrow h = 2.$

(iii) Using only the information readable on the solution table and data for only the first three months, determine n^* for the first three months. [15%]

ANSWER: $n^* = 3$

s_t	200	200	200	200
d_t	??	29	30	X
$Y_t(1)$	200	258	378	$378+6X$
$Y_t(2)$		400	460	$460+4X$
$Y_t(3)$			458	$458+2X$
$Y_t(4)$				578
z_t	200	258	378	
δ_t	(1)	(1,0)	(1,0,0)	

(iv) You are concerned about how demand in month 4, d_4, might impact your answer to part (iii). Either find ranges of values for d_4 for which the answer to part (iii) would change or else show that the answer to part (iii) would not change whatever the value of d_4. [15%]

ANSWER: One can find a value of d_4 where the answer to part (iii) would change. For example, if $X = 50$, $\delta_4 = (1,0,1,0)$. Therefore, it is not possible to show that the answer to part (iii) would not change whatever the value of d_4.

We will thus find ranges of values for d_4 for which the answer to part (iii) would change. Note that the answer to part (iii) would not change either if $\delta_4 = (1,0,0,1)$ or $\delta_4 = (1,0,0,0)$. The first is possible, if

$$578 \leq min(378+6X, 460+4X, 458+2X) \quad \Rightarrow \quad X \geq max(200/6, 118/4, 60) = 60.$$

The second is possible, if

$$378+6X \leq min(578, 460+4X, 458+2X) \quad \Rightarrow \quad X \leq min(200/6, 82/4, 20) = 20.$$

That is, when $20 < X < 60$, the answer to part (iii) would change.

(v) Are there values of d_4 such that n^* would remain fixed regardless of the values of d_5? If so, what are they? If not, why not? (Please note that n^* here is not necessarily equal to the values in the previous parts.) [15%]

ANSWER: In order for n^* to remain fixed regardless of values of d_5, we should have a $\delta_t = (1, **, **, 1)$. That is,

$$578 \leq min(378+6X, 460+4X, 458+2X) \quad \Rightarrow \quad X \geq max(200/6, 118/4, 60) = 60.$$

(b) A manufacturer of small electric motors uses an automatic milling machine to produce slots in motor shafts. A batch of shafts is run and then checked. All shafts that do not meet required dimensional tolerances are discarded. At the beginning of each new batch, the milling machine needs to be adjusted, because the machine's cutter head wears slightly during batch production.

The manufacturer would like to predict the number of defective shafts per batch as a function of the size of the batch. To this end, the engineering department has gathered data on 30 batches; the summary statistics of the data is provided below:

	Batch size	**Number of defective shafts**
Mimimum	100	5
Average	252.5	44.8
Maximum	400	112

The engineering department estimated a linear regression model to predict the number of defective shafts per batch as a function of the batch size. The output is reproduced on the next page.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.98
R Square	0.95
Adjusted R Square	0.95
Standard Error	7.56
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	32744.46	32744.46	572.9	0.00
Residual	28	1600.34	57.16		
Total	29	34344.80			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-47.9	4.11	-11.65	0.00	-56.32	-39.48
Batch size	0.37	0.02	23.94	0.00	0.34	0.40

(i) What is the regression equation produced by the linear regression model? [5%]

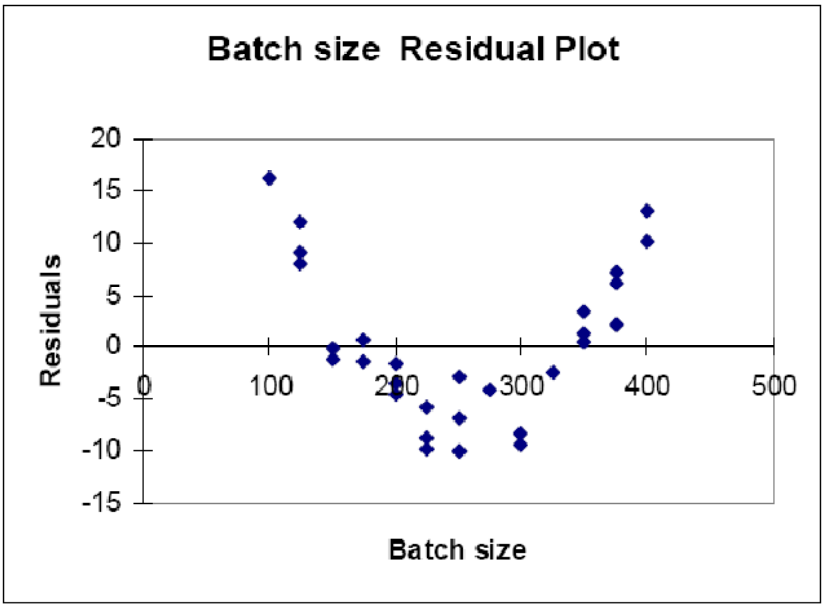
ANSWER: Predicted number of defective shafts = -47.9 + 0.37 x batch size.

(ii) According to this model, what is the predicted number of defective shafts per batch if the size of the batch is 50? What can you say about the validity of this prediction? [10%]

ANSWER: Predictions made using an estimated regression function may have little or no validity for values of the independent variables that are substantially different from those represented in the sample. For example, in this case, batch size 50 < the minimum batch size observed in the data, and the predicted number of defective shafts = -47.9 + 0.37 x 50 = -29.4, is quite unrealistic.

Hence, be cautious when you want to make predictions for a value that is far away from given observations for the independent variable.

(iii) Look at the residual plot that is produced with the regression output. What can you say about the model? Do you recommend using this model? [20%]



ANSWER: The R^2 value is 0.95, which is very close to one, and so, our linear regression model explains a very high percentage of the variation. The t-statistics are also large, indicating that the coefficients are statistically significant. The linear model is a good fit to the data, but looking at the residual plot there is evidence that there is a convex (quadratic) dependency of the residuals on the batch size.

END OF PAPER