

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
Summer 2025

This paper is also taken for the relevant examination for the
Associateship of the Royal College of Science

Probability and Statistics

Date: Friday, May 9, 2025

Time: Start time 14:00 – End time 17:00 (BST)

Time Allowed: 3 hours

This paper has 6 Questions.

Please Answer Each Question in a Separate Answer Booklet

This is a closed book examination.

Candidates should start their solutions to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Allow margins for marking.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO DO SO

Please remember to justify all your answers and state carefully which results from the lectures you apply in your proofs.

Statistical tables are provided on the last pages of the exam paper.

1. (a) (i) Briefly describe the probability space $(\Omega, \mathcal{F}, P)$ for rolling a fair six-sided die. (3 marks)
- (ii) A fair n -sided die is rolled multiple times, with each roll being independent of the others. Let $X_1, X_2, \dots, X_n$ represent the outcomes of n rolls, where each $X_i \in \{1, 2, \dots, n\}$. We say a *match* occurs on the i -th roll if $X_i = i$, i.e. the number rolled on the i -th roll equals the roll number itself. Let M be the random variable that counts the number of matches (the number of rolls for which $X_i = i$ for $i = 1, 2, \dots, n$).
- (a) Determine a suitable distribution for the random variable X_i (the outcome of the i -th roll of the die) and the random variable M (the number of matches). (3 marks)
- (b) What is the probability p_n that at least one match occurs in the first n rolls? (3 marks)

[Question 1 continues on the next page.]

- (b) You can travel from Oxford to London Paddington via three distinct rail routes as shown below:

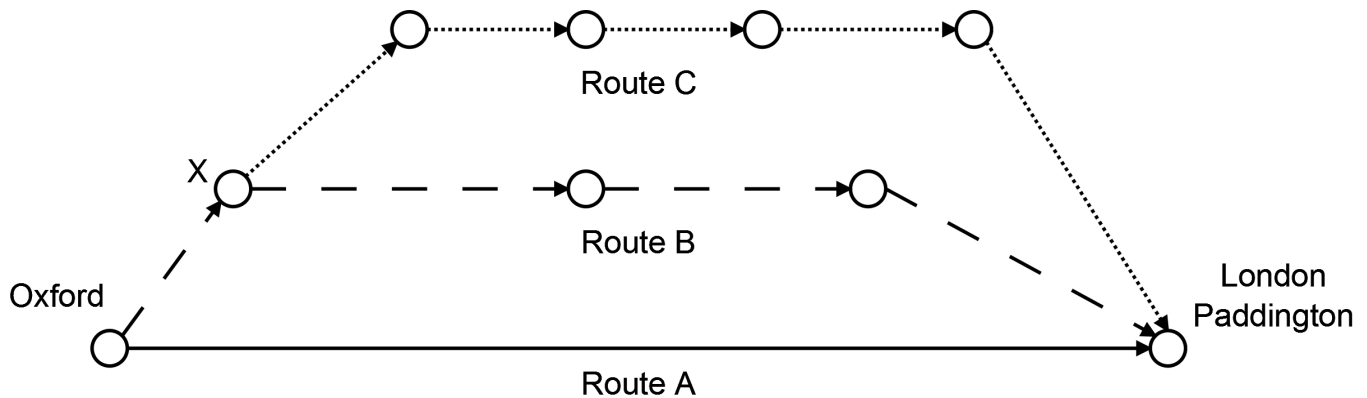


Figure 1: Diagrams showing train routes A, B and C.

Route A: The straight solid line at the bottom directly connecting Oxford to Paddington.

Route B: The dashed line that travels through intermediate stations.

Route C: The route following the dashed line from Oxford to X and the dotted line from X to Paddington.

The following table summarises the information available for routes A, B and C in a certain time period.

Route	A	B	C
Number of services	100	60	40
Number of services being late	4	6	2
Total number of passengers	30	20	10

Assuming that you are waiting for a guest to arrive at London Paddington from Oxford, answer the following questions and state clearly any additional assumptions you make. You may leave your answers in terms of fractions that do not need to be simplified further.

- What is the probability that the guest will be arriving late at London Paddington?
(5 marks)
- The guest boarded on a train at Oxford but has informed you that his train has been delayed and will be arriving late. As he is new to the country he is unaware of which route he has taken. What is the probability that the guest took a train that went via route A?
(2 marks)
- It turns out that a signal failure has occurred at junction X causing all trains passing through the junction to be late. What is now the probability that the guest took route A?
(4 marks)

(Total: 20 marks)

2. (a) You work at a cybersecurity company, and you are responsible for designing secure passwords for client systems. A specific client has requested that you create passwords as a string of 10 characters such that:

- * 5 of the characters are letters (upper and lower case, i.e. 52 characters) with repetitions allowed,
- * 3 are one-digit numbers $\{0,1,2,3,4,5,6,7,8,9\}$ with repetitions allowed, and
- * 2 are distinct symbols from the list of 5 symbols: $\{!, @, \#, \$, \&\}$.

(i) How many possible passwords are there? [You may leave your answer in terms of factorials or binomial coefficients that do not need to be simplified further.]

(5 marks)

(ii) With all locations for symbols, letters, or numbers in your 10-character password being equally likely, what is the probability that the two symbols are next to each other?

(4 marks)

(b) Consider jointly continuous random variables X and Y with joint probability density function (pdf) given by:

$$f_{X,Y}(x,y) = \begin{cases} c(x^2 + y^2), & \text{for } 0 \leq y < x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Show that, for $c = 3$, $f_{X,Y}$ is a valid probability density function. (3 marks)

(ii) Find the marginal density of X . (3 marks)

(iii) Find the conditional probability density function of Y given X . (2 marks)

(iv) Using your answer in part (iii), comment on whether the variables X and Y are independent (no calculations needed). Suggest an alternative way of checking independence. (3 marks)

(Total: 20 marks)

3. (a) (i) Let U be a Uniform random variable on the interval $(0, 1)$, and let $\lambda > 0$. Find the distribution of the random variable $-(\log U)/\lambda$.
(4 marks)
- (ii) Consider a sequence of $n \in \mathbb{N}$ independent random variables $X_1, \dots, X_n$, where each random variable follows the Uniform distribution on the interval $(0, 1)$, i.e. $X_i \sim U(0, 1), i = 1, \dots, n$. Find the cumulative distribution function of $Y = \max \{X_1, \dots, X_n\}$.
(4 marks)
- (b) (i) Let X be a standard normal variable. Compute the moment generating function.
(3 marks)
- (ii) Suppose $X_1, X_2, \dots, X_n$ are independent standard normal random variables. Using the result from part (i), find the distribution of the sum $S = X_1 + X_2 + \dots + X_n$, and state clearly the parameters.
(4 marks)
- (c) (i) Let U and V be two random variables such that $\mathbb{E}[U^k] = \mathbb{E}[V^k]$ for all non-negative integers k . Show that U and V have the same distribution and state clearly any additional assumptions you make. [You may assume that the expectations exist and are finite, i.e. $\mathbb{E}[U^k] < \infty$ and $\mathbb{E}[V^k] < \infty$ for all non-negative integers k .]
(4 marks)
- (ii) Does the conclusion in part (i) depend on whether U and V are continuous, discrete, or mixed random variables?
(1 mark)

(Total: 20 marks)

4. (a) Suppose $X_1, X_2, \dots, X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma^2)$ for $i = 1, 2, \dots, n$. Furthermore, suppose that $\mu_i \neq \mu_j$, for $i \neq j$ and $i, j \in \{1, 2, \dots, n\}$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \bar{\mu} = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

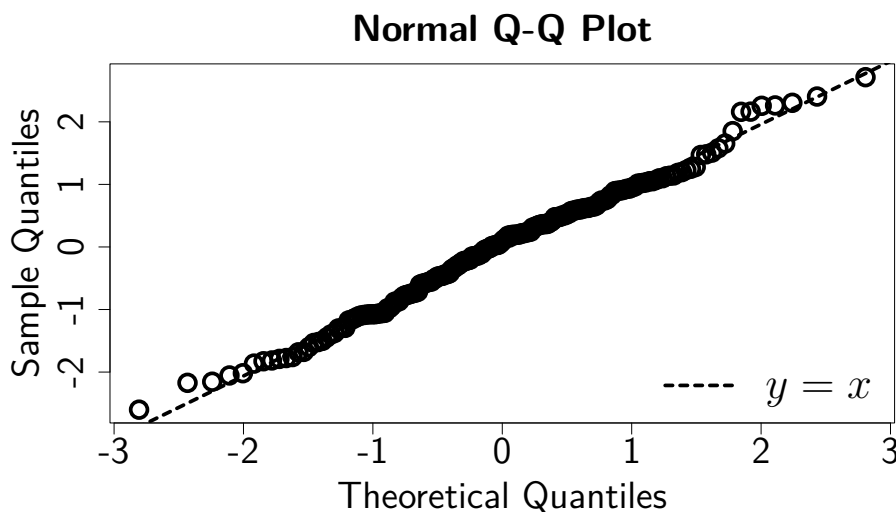
- (i) Prove $\text{Var}[\bar{X}] = \frac{\sigma^2}{n}$. (2 marks)
- (ii) Compute $E[S^2]$, clearly stating any results used and providing full details. (5 marks)

- (b) The heights of the Imperial College football team are measured as $x_1, x_2, \dots, x_n$, where $n = 12$. The sample mean and sample variance of these observations are computed as $\bar{x} = 181\text{cm}$ and $s^2 = 3\text{cm}^2$, respectively. Let's assume that the heights are observations of the independent random variables $X_1, X_2, \dots, X_n$ which follow a normal distribution with mean μ and variance σ^2 , where μ and σ^2 are both unknown.

- (i) Compute a 90% confidence interval for the unknown mean, μ , with full justification. Tables 1-4 on pages 10-13 may be helpful. (4 marks)
- (ii) Compute a 95% confidence interval for the unknown variance, σ^2 , with full justification. Tables 1-4 on pages 10-13 may be helpful. (4 marks)

- (c) Suppose a 99% confidence interval for a parameter θ is computed as $(3, 7)$. Provide a careful interpretation of this confidence interval in relation to the true value of θ . (3 marks)

- (d) Suppose a sample of $n = 200$ observations are recorded. A statistician wants to test the assumption that the data are observations of a normal random variable, so they plot the Q-Q plot below. Interpret the Q-Q plot below, commenting on whether the plot provides evidence that the data are normally distributed or not. (2 marks)



(Total: 20 marks)

5. (a) Suppose that the $n = 13$ observations are recorded:

$\{0.9, 5.3, 4.5, 7.5, 4.9, 2.0, 3.0, 1.1, 3.4, 9.8, 1.7, 0.8, 3.9\}$

- (i) Compute the median of this dataset. (1 mark)
- (ii) Compute the upper and lower quartiles of this dataset. (2 marks)
- (iii) Use Tukey's criterion to identify any outliers in this dataset. (1 mark)
- (iv) Which type of plot would be best to visualise the median, lower quartile, upper quartile and any outliers of this dataset? (1 mark)

- (b) Suppose that a statistician decides to compare the average heights of the students in the Imperial rugby team and basketball team. The teams consist of different students; no student in the university is in both teams. The heights of the students in the rugby team are assumed to follow a $N(\mu_1, \sigma_1^2)$ distribution, and the heights of the students in the basketball team are assumed to follow a $N(\mu_2, \sigma_2^2)$ distribution, where $\sigma_1^2 = 10$ and $\sigma_2^2 = 4$ are known, but μ_1 and μ_2 are unknown. Carefully stating any assumptions and results used, test the hypothesis that $\mu_1 = \mu_2$ at the significance threshold $\alpha = 0.01$, given that the heights of the rugby team are measured as $x_1, x_2, \dots, x_n$ and the heights of the basketball team are measured as $y_1, y_2, \dots, y_m$, such that the units for height are in cm and

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = 182, & n &= 15, \\ \bar{y} &= \frac{1}{m} \sum_{i=1}^m y_i = 185, & m &= 12,\end{aligned}$$

As part of the calculation, ensure that you compute the p -value for the hypothesis test. Note that Tables 1-4 on pages 10-13 may be helpful. (7 marks)

- (c) Give definitions for Type I and Type II errors of a hypothesis test. Explain how these errors depend on the significance threshold α and discuss if increasing the value of α would either increase or decrease the probabilities of these errors occurring. (4 marks)
- (d) Suppose that a medical research lab is testing for the association of different genetic variants with a particular disease. Each genetic variant is investigated in a separate experiment which results in a p -value, and a significance threshold of $\alpha = 0.05$ is specified in advance for each experiment. Out of the 10 genetic variants that were tested, the five most significant p -values were

$\{0.04, 0.007, 0.002, 0.024, 0.09\}$.

Given the experimental setup described, which of these p -values should be declared to be significant? Provide justification and state any results used. (4 marks)

(Total: 20 marks)

6. (a) Suppose the data $x_1, x_2, \dots, x_n$ are measurements of the journey time of a particular London bus route over several days. A statistician decides to model these as independent observations from a continuous uniform distribution on the interval $[\theta_1, \theta_2]$, where $\theta_1, \theta_2 \in (0, \infty)$ are unknown. Compute the maximum likelihood estimates of θ_1 and θ_2 , given the observed data, and write down expressions for their associated maximum likelihood estimators. (6 marks)

- (b) Given the n pairs of observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ for quantities X and Y , the simple linear regression model is given by

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i \in \{1, 2, \dots, n\},$$

where the $\epsilon_i \sim N(0, \sigma^2)$, $i \in \{1, 2, \dots, n\}$, are independent and identically distributed with σ^2 unknown. One can obtain maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the parameters β_0 and β_1 , respectively, as

$$\hat{\beta}_0 = \hat{\beta}_0(\mathbf{x}, \mathbf{y}) = \bar{y} - \left(\frac{S_{xy}}{S_{xx}} \right) \bar{x}, \quad \hat{\beta}_1 = \hat{\beta}_1(\mathbf{x}, \mathbf{y}) = \frac{S_{xy}}{S_{xx}},$$

where $\bar{x}$, $\bar{y}$, S_{xx} , S_{yy} and S_{xy} are defined by

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, & S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2, \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i, & S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2, \\ S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \end{aligned}$$

- (i) Show that one can equivalently write S_{xy} as $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})y_i$. (1 mark)
- (ii) Use the expression in (i) to conclude that the maximum likelihood estimator of $\hat{\beta}_1$ is $\hat{\beta}_1 = \hat{\beta}_1(\mathbf{x}, \mathbf{Y}) = \sum_{i=1}^n (x_i - \bar{x})Y_i / S_{xx}$. (1 mark)
- (iii) Use the expression in (i) to conclude that the maximum likelihood estimator of $\hat{\beta}_0$ is $\hat{\beta}_0 = \hat{\beta}_0(\mathbf{x}, \mathbf{Y}) = \sum_{i=1}^n (1/n - (x_i - \bar{x})\bar{x}/S_{xx})Y_i$, and compute $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$. (4 marks)
- (iv) Briefly describe a real-world example where linear regression can be used to investigate a relationship between two quantities X and Y . If there is no clear linear relationship between X and Y , describe how linear regression can still be used to investigate a relationship between these two quantities. (4 marks)

- (c) Suppose a professor is teaching a class with 250 students. After a class test, all 250 scripts are randomly divided among 10 assistant markers, with each assistant marker receiving 25 scripts to mark. The day after the test, the first assistant marker returns their marked scripts to the professor. Each script has a test score which is an integer between 0 and 20 (inclusive). Concisely describe how the professor could use the bootstrap procedure with the data from the first batch of 25 marked scripts to construct a 95% confidence interval for the median test score of the class, and give an expression for this confidence interval. (4 marks)

(Total: 20 marks)

Name of distribution of X	$\text{Im}X$	Parameters	p.m.f. $P(X = x)/\text{p.d.f } f_X(x)$
Discrete distributions			
Bernoulli: $X \sim \text{Bern}(p)$	$\{0, 1\}$	$p \in (0, 1)$	$p^x(1-p)^{1-x}$
Binomial: $X \sim \text{Bin}(n, p)$	$\{0, 1, \dots, n\}$	$n \in \mathbb{N}, p \in (0, 1)$	$\binom{n}{x} p^x (1-p)^{n-x}$
Hypergeometric: $X \sim \text{HGeom}(N, K, n)$	$\{0, 1, \dots, \min(n, K)\}$	$N \in \mathbb{N} \cup \{0\},$ $K, n \in \{0, 1, \dots, N\}$	$\frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}},$ for $x \in \{0, 1, \dots, K\},$ $n - x \in \{0, 1, \dots, N - K\}$
Poisson: $X \sim \text{Poi}(\lambda)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$\lambda > 0$	$\frac{\lambda^x}{x!} e^{-\lambda}$
Geometric (1): $X \sim \text{Geom}(p)$	$\mathbb{N} = \{1, 2, \dots\}$	$p \in (0, 1)$	$(1-p)^{x-1} p$
Geometric (2): $X \sim \text{Geom}(p)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$p \in (0, 1)$	$(1-p)^x p$
Negative binomial: $X \sim \text{NBin}(r, p)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$r \in \mathbb{N}, p \in (0, 1)$	$\binom{x+r-1}{r-1} p^r (1-p)^x$
Continuous distributions			
Uniform: $X \sim \text{U}(a, b)$	(a, b)	$a, b \in \mathbb{R}, a < b$	$\frac{1}{b-a}$
Exponential: $X \sim \text{Exp}(\lambda)$	$(0, \infty)$	$\lambda > 0$	$\lambda e^{-\lambda x}$
Gamma: $X \sim \text{Gamma}(\alpha, \beta)$	$(0, \infty)$	$\alpha, \beta > 0$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$ where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
Chi-squared: $X \sim \chi^2(n)$ $(X \sim \chi_n^2)$	$(0, \infty)$	$n \in \mathbb{N}$	$\frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{n/2-1} e^{-x/2}$
Beta: $X \sim \text{Beta}(\alpha, \beta)$	$(0, 1)$	$\alpha, \beta > 0$	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$ where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$ $= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
Normal: $X \sim \text{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Student-t: $X \sim \text{Student}(\nu)$	$\mathbb{R}$	$\nu > 0$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

Table 1: $P(Z < z)$ where $Z \sim N(0, 1)$ for values of z between 1.00 and 3.99. One reads the table as $P(Z < z) = p_{r,c}$, where $z = z_r + z_c$ and z_r are the first two digits of z shown in row r , z_c is the third digit of z shown in column c , and $p_{r,c}$ is the (r, c) th entry. For example, $P(Z < 1.02) = 0.8461$, looking in row labelled 1.0 and column labelled 0.02.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2: Values of t for $P(T < t)$, where T has Student's t -distribution with ν degrees of freedom, for $\nu \in \{1, 2, \dots, 30\}$. In this table, $P(T < t_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $t_{\nu,p}$ is the corresponding entry in the table. For example, $t_{3,0.60} = 0.277$, looking in row labelled 3 and column labelled 0.60.

ν	0.60	0.667	0.75	0.80	0.875	0.90	0.95	0.975	0.99	0.995	0.999
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385

Table 3: Values of x for $P(X < x)$, where $X \sim \chi^2(\nu)$, a chi-squared distribution with ν degrees of freedom. In this table, $P(T < \chi^2_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $\chi^2_{\nu,p}$ is the corresponding entry in the table. For example, $\chi^2_{3,0.005} = 0.072$, looking in row labelled 3 and column labelled 0.005.

ν	0.001	0.005	0.01	0.025	0.05	0.1	0.125	0.2	0.25	0.333	0.5
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335

Table 4: Values of x for $P(X < x)$, where $X \sim \chi^2(\nu)$, a chi-squared distribution with ν degrees of freedom. In this table, $P(T < \chi^2_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $\chi^2_{\nu,p}$ is the corresponding entry in the table. For example, $\chi^2_{3,0.6} = 2.946$, looking in row labelled 3 and column labelled 0.6.

ν	0.6	0.66	0.75	0.80	0.87	0.9	0.95	0.975	0.99	0.995	0.999
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661

1. (a) (i) The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for rolling a fair six-sided die can be described as follows:

meth seen ↓

Sample Space (Ω): The sample space consists of all possible outcomes of rolling the die. Since the die has six faces, we have:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

1, A

Event space ($\mathcal{F}$): The event space is the collection of all subsets of Ω , which represents all possible events, including the empty set, the entire sample space, and each individual outcome:

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \dots, \{6\}, \{1, 2\}, \{1, 3\}, \dots, \{1, 6\}, \dots, \{1, 2, \dots, 6\}\}.$$

1, B

Probability Measure (P): The probability measure is a mapping $P : \mathcal{F} \rightarrow \mathbb{R}$, meaning that we assign probabilities to each event in the event space and these probabilities must be real numbers between 0 and 1 inclusive. For a fair die, each face has an equal probability of appearing, so:

$$P(\{i\}) = \frac{1}{6}, \quad \text{for each } i \in \{1, 2, 3, 4, 5, 6\}.$$

For any event $A \in \mathcal{F}$, the probability of A is given by:

$$P(A) = \sum_{i \in A} P(\{i\}) = \frac{|A|}{6},$$

where $|A|$ is the number of elements in A .

1, B

- (ii) (a) Since each outcome of a roll is equally likely from 1 to n , this would be a *Discrete Uniform distribution* over the set $\{1, 2, \dots, n\}$.

A match occurs on the i -th roll if $X_i = i$, which can be viewed as a Bernoulli trial with success probability $\frac{1}{n}$ (because the probability of rolling a i on the i -th roll is $\frac{1}{n}$). Since M is the sum of n independent Bernoulli trials (one for each roll), the total number of matches M follows a *Binomial distribution*.

3, B

[2 marks for correct distributions, 1 mark for justification]

unseen ↓

- (b) To calculate the probability that at least one match occurs in the first n rolls of a fair n -sided die, we use the principle of inclusion-exclusion. The probability p_n can be expressed as:

$$p_n = 1 - P(\text{no matches in the first } n \text{ rolls}).$$

There are $n - 1$ outcomes for the i -th roll that are not i . Thus, the probability of no match on a single roll is $\frac{n-1}{n}$.

Since the rolls are independent, the probability of no match on all n rolls is:

$$P(\text{no matches}) = \left(\frac{n-1}{n}\right)^n.$$

Hence, the probability p_n that at least one match occurs in the first n rolls is given by:

$$p_n = 1 - \left(\frac{n-1}{n}\right)^n.$$

[2 marks for estimating the probabilities (probability of no matches, probability of at least one match), 1 mark for justification]

3, C

- (b) (i) Assume that the guest is selected randomly from the 60 passengers. Let L denote the event that a train is late and let A, B and C denote the events that a passenger takes routes A, B and C, respectively. Therefore, an estimate of the conditional probability of a train being late for a given route is:

$$P(L|A) = \frac{4}{100} = \frac{1}{25}, \quad P(L|B) = \frac{6}{60} = \frac{1}{10}, \quad P(L|C) = \frac{2}{40} = \frac{1}{20}.$$

The total number of passengers is $30 + 20 + 10 = 60$.

An estimate of the probability that a passenger takes each route is:

$$P(A) = \frac{30}{60} = \frac{1}{2}, \quad P(B) = \frac{20}{60} = \frac{1}{3}, \quad P(C) = \frac{10}{60} = \frac{1}{6}.$$

Using the Law of Total Probability (LTP):

$$\begin{aligned} P(L) &= P(L|A)P(A) + P(L|B)P(B) + P(L|C)P(C) \\ &= \frac{1}{25} \cdot \frac{1}{2} + \frac{1}{10} \cdot \frac{1}{3} + \frac{1}{20} \cdot \frac{1}{6} \\ &= \frac{1}{50} + \frac{1}{30} + \frac{1}{120} \\ &= \frac{12}{600} + \frac{20}{600} + \frac{5}{600} = \frac{37}{600}. \end{aligned}$$

[1 mark for the assumption, 1 mark for introducing the notation and reading off the probabilities, 1 mark for mentioning the Law of Total Probability (LTP), 2 marks for the calculations.]

5, A

- (ii) Using Bayes' Theorem:

$$P(A|L) = \frac{P(L|A)P(A)}{P(L)} = \frac{\frac{1}{25} \cdot \frac{1}{2}}{\frac{37}{600}} = \frac{600}{1850} = \frac{12}{37}.$$

[1 mark for mentioning the Bayes' Theorem, 1 mark for the calculations.]

- (iii) The question informs you that trains passing through junction X will be late. From the diagram, this means that routes B and C will be affected. Hence, the conditional probabilities have now changed:

2, A

unseen ↓

$$P(L|B) = P(L|C) = 1$$

Now, we re-calculate the probability that the train is late, using the Law of Total Probability:

$$P(L) = P(L|A)P(A) + P(B) + P(C) = \frac{1}{25} \cdot \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = \frac{78}{150} = \frac{13}{25}.$$

Hence, re-apply Bayes' Theorem:

$$P(A|L) = \frac{P(L|A)P(A)}{P(L)} = \frac{\frac{1}{25} \cdot \frac{1}{2}}{\frac{13}{25}} = \frac{1}{26}.$$

[1 mark for justifying the conditional probabilities (conditions on B and C), 1 mark for mentioning both the Law of Total Probability (LTP) and the Bayes' theorem, 2 marks for the calculations.]

4, D

2. (a) (i) Choosing the positions for letters, numbers, and symbols:

meth seen ↓

First, we choose 2 positions for the symbols. The number of ways to choose 2 positions out of 10 is given by the binomial coefficient $\binom{10}{2}$.

After selecting the 2 positions for symbols, 8 positions remain for the numbers and symbols.

Next, we choose positions for the letters. The number of ways to choose 5 positions out of these remaining 8 is $\binom{8}{5}$.

The remaining 3 positions must be for the numbers, and there is only 1 way to assign these remaining positions to the numbers, $\binom{3}{3} = 1$.

Choosing the characters for each category:

The number of ways to choose 2 distinct symbols is given by $5 \cdot 4 = 20$ (since the order of the symbols matters).

For the 3 number positions, each number can be one of 10 possibilities (0-9), so there are 10^3 possible ways to choose the numbers.

For the 5 letter positions, each letter can be one of 52 possibilities (uppercase or lowercase), so there are 52^5 possible ways to choose the letters.

So, the number of possible passwords is $\binom{10}{2} \cdot \binom{8}{5} \cdot 20 \cdot 10^3 \cdot 52^5$.

5, B

[1 mark for choosing the positions for letters, numbers and symbols, 1 mark for choosing the characters for each category, 2 marks for justification, 1 mark for the calculations]

unseen ↓

- (ii)
- Choose the positions of the two adjacent symbols: There are 9 ways to place two adjacent symbols.
 - Arrange the two symbols: There are $5 \cdot 4 = 20$ ways to choose two distinct symbols.
 - Choose the locations of the letters: $\binom{8}{5}$.
 - Choose the 5 letters (with replacement): 52^5 .
 - Choose the locations of the digits: $\binom{3}{3} = 1$.
 - Choose the digits: 10^3 .

Thus, the total number of favorable outcomes (the number of passwords where the two symbols are adjacent) is $9 \cdot \binom{8}{5} \cdot 20 \cdot 10^3 \cdot 52^5$, and the probability that the two symbols are next to each other is given by:

$$P(\text{two adjacent symbols}) = \frac{\text{Favorable outcomes}}{\text{Total outcomes}} = \frac{9 \cdot \binom{8}{5} \cdot 20 \cdot 10^3 \cdot 52^5}{\binom{10}{2} \cdot \binom{8}{5} \cdot 20 \cdot 10^3 \cdot 52^5} = \frac{9}{\binom{10}{2}} = \frac{2}{10} = \frac{1}{5}.$$

[1 mark for identifying the total number of possible passwords, 1 mark for the calculations, 1 mark for justification]

Alternatively, full marks will be given for the ratio of the number of ways to place the 2 symbols adjacently, 9, to the total number of ways to choose the 2 positions, $\binom{10}{2}$, on the grounds that the other components of the password are irrelevant.]

4, D

(b) (i) We note that:

meth seen ↓

$$\begin{aligned}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy dx &= \int_0^1 \int_0^x c(x^2 + y^2) dy dx \\&= c \int_0^1 \int_0^x (x^2 + y^2) dy dx \\&= c \int_0^1 \left[x^2 \int_0^x 1 dy + \int_0^x y^2 dy \right] dx \\&= c \int_0^1 x^2 \cdot y \Big|_{y=0}^{y=x} + \frac{y^3}{3} \Big|_{y=0}^{y=x} dx \\&= c \int_0^1 \left(x^3 + \frac{x^3}{3} \right) dx = c \int_0^1 \frac{4x^3}{3} dx \\&= c \cdot \frac{4}{3} \int_0^1 x^3 dx = c \cdot \frac{4}{3} \cdot \frac{x^4}{4} \Big|_{x=0}^1 = c \cdot \frac{4}{3} \cdot \frac{1}{4} = c \cdot \frac{1}{3} \stackrel{!}{=} 1,\end{aligned}$$

if and only if $c = 3$. We note that, for $c = 3$, the density is nonnegative for all x, y and integrates to 1.

[1 mark for integral equal to 1, 1 mark for the calculations, 1 mark for checking non-negativity]

3, A

(ii) The marginal density of X is given by:

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \\&= \int_0^x 3(x^2 + y^2) dy \\&= 3 \int_0^x x^2 dy + 3 \int_0^x y^2 dy \\&= 3x^2 \int_0^x 1 dy + 3 \int_0^x y^2 dy \\&= 3x^3 + 3 \cdot \frac{y^3}{3} \Big|_{y=0}^{y=x} \\&= 3x^3 + 3 \cdot \frac{x^3}{3} = 4x^3.\end{aligned}$$

Hence $f_X(x) = 4x^3$ for $0 \leq x \leq 1$ and 0 otherwise.

[1 mark for method/integral limits, 1 mark for obtaining the correct marginal density, 1 mark for stating clearly the range]

3, A

(iii) The conditional probability density function of Y given X is defined as:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},$$

where $f_{X,Y}(x,y)$ is the joint density function and $f_X(x)$ is the marginal density of X . Hence, the conditional density function is given by:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{3(x^2 + y^2)}{4x^3}, \quad 0 \leq y < x \leq 1.$$

2, A

- (iv) We observe that the conditional density $f_{Y|X}(y|x)$ is not the same as a marginal density $f_Y(y)$, which would not involve x . This implies that X and Y are not independent. Alternatively, we could estimate the marginal density of Y , and check whether the product of the marginal densities $f_X(x)f_Y(y)$ gives us the joint probability density function $f_{X,Y}(x,y)$.
[1 mark for showing non-independence, 1 mark for justification, 1 mark for suggesting an alternative way]

3, C

3. (a) (i) Let $X = -\frac{\log U}{\lambda}$, where U is a uniform random variable on $(0, 1)$. By definition, the cumulative distribution function (cdf) of X , $F_X(x)$, is given by:

meth seen ↓

$$F_X(x) = P(X \leq x) = P\left(-\frac{\log U}{\lambda} \leq x\right) = P(U \geq e^{-\lambda x}) = 1 - P(U < e^{-\lambda x}).$$

Since $U \sim U(0, 1)$, we have $P(U \leq u) = \int_0^u 1 dt = u$ for $u \in [0, 1]$. Therefore, $P(U < e^{-\lambda x}) = e^{-\lambda x}$, and the cdf of X is $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.

Now, we differentiate $F_X(x)$ with respect to x to get the pdf of X :

$$f_X(x) = \frac{d}{dx} (1 - e^{-\lambda x}) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0.$$

Therefore, $X \sim \text{Exp}(\lambda)$.

[1 mark for deriving the cdf, 1 mark for deriving the pdf, 1 mark for recognising the exponential distribution, 1 mark for justification. If students recognise the distribution without deriving the pdf, then full marks will be awarded provided the reasoning is clearly explained.]

4, A

- (ii) By definition, the cumulative distribution function (cdf) of $Y = \max\{X_1, X_2, \dots, X_n\}$, $F_Y(y)$, is given by:

$$F_Y(y) = P(Y \leq y) = P(\max\{X_1, X_2, \dots, X_n\} \leq y) = P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y).$$

Since $X_1, X_2, \dots, X_n$ are independent random variables, we can write:

$$= P(X_1 \leq y) \cdot P(X_2 \leq y) \cdots P(X_n \leq y).$$

Since $X_i \sim U(0, 1)$, we have $P(X_i \leq y) = \int_0^y 1 dx = y$ for $y \in [0, 1]$. Hence, the cdf of Y is $P(Y \leq y) = y \cdot y \cdots y = y^n$ for $y \in [0, 1]$.

For $y < 0$, we have $P(Y \leq y) = 0$, and for $y > 1$, $P(Y \leq y) = 1$. Hence, the cdf of Y is given by:

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0, \\ y^n & \text{if } 0 \leq y \leq 1, \\ 1 & \text{if } y > 1. \end{cases}$$

[1 mark for deriving the cdf, 2 marks for justification (definition, independence), 1 mark for stating the full piecewise form of cdf]

4, A

- (b) (i) The moment generating function $M_X(t)$ is defined as $M_X(t) = \mathbb{E}[e^{tX}]$. Thus, using the Law of the Unconscious Statistician (LOTUS), the moment generating function of a standard normal variable becomes:

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} e^{\frac{t^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx = e^{\frac{t^2}{2}}, \end{aligned}$$

since the latter integral is equal to 1 since it is the integral of a $N(t, 1)$ density function.

Thus, the moment generating function of a standard normal random variable X is:

$$M_X(t) = e^{\frac{t^2}{2}}, \quad \text{for all } t \in \mathbb{R}.$$

[1 mark for mentioning LOTUS, 1 mark for method/integration, 1 mark for the calculations]

3, B

- (ii) We have seen in the lectures, that, since $S = X_1 + X_2 + \dots + X_n$, and the X_i are independent, the moment generating function (mgf) of S is the product of the mgfs of $X_1, X_2, \dots, X_n$:

$$M_S(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t).$$

Thus, the mgf of S becomes $M_S(t) = (M_{X_i}(t))^n = (e^{t^2/2})^n = e^{nt^2/2}$.

3, C

We recognise this is the mgf of a normal distribution with mean 0 and variance n . Hence, $S \sim N(0, n)$.

1, B

[2 marks for justification (independence, product of mgfs), 1 mark for the calculations, 1 mark for recognising the normal distribution]

unseen ↓

- (c) (i) The moment generating functions (mgfs) $M_U(t)$ $M_V(t)$ are defined as:

$$M_U(t) = \mathbb{E}[e^{tU}], \quad M_V(t) = \mathbb{E}[e^{tV}].$$

We can expand the exponential functions e^{tU} and e^{tV} as their respective Taylor series at $t = 0$:

$$e^{tU} = 1 + tU + \frac{t^2}{2!}U^2 + \frac{t^3}{3!}U^3 + \dots$$

$$e^{tV} = 1 + tV + \frac{t^2}{2!}V^2 + \frac{t^3}{3!}V^3 + \dots$$

Now, we take the expectation of both series:

$$\mathbb{E}[e^{tU}] = \mathbb{E}\left[1 + tU + \frac{t^2}{2!}U^2 + \frac{t^3}{3!}U^3 + \dots\right] = \mathbb{E}[1] + t\mathbb{E}[U] + \frac{t^2}{2!}\mathbb{E}[U^2] + \frac{t^3}{3!}\mathbb{E}[U^3] + \dots$$

Similarly, $\mathbb{E}[e^{tV}] = \mathbb{E}[1] + t\mathbb{E}[V] + \frac{t^2}{2!}\mathbb{E}[V^2] + \frac{t^3}{3!}\mathbb{E}[V^3] + \dots$

We are given that $\mathbb{E}[U^k] = \mathbb{E}[V^k]$ for all non-negative integers k . Thus, comparing the Taylor series for $\mathbb{E}[e^{tU}]$ and $\mathbb{E}[e^{tV}]$, we see that:

$$\mathbb{E}[e^{tU}] = \mathbb{E}[e^{tV}].$$

Since $\mathbb{E}[e^{tU}] = \mathbb{E}[e^{tV}]$ for all t in the domain where the series converge, we conclude that U and V have the same moment generating function, $M_U(t) = M_V(t)$ for all t , and hence, since the mgfs of U and V are equal, and the mgf uniquely determines the distribution of a random variable (given that it exists in some neighborhood of zero), we conclude that U and V have the same distribution.

4, D

- (ii) No, the assumption of whether U and V are discrete, continuous or mixed random variables does not change the result. The equality of moments ($\mathbb{E}[U^k] = \mathbb{E}[V^k]$ for all $k \geq 0$) guarantees that U and V have the same distribution, regardless of their type.

1, B

4. (a) (i)

seen ↓

$$\begin{aligned}
 \text{Var}[\bar{X}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\
 &= \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] && \text{(by a property of the variance)} \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] && \text{(since the random variables are independent)} \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\
 &= \frac{n\sigma^2}{n^2} \\
 \Rightarrow \text{Var}[\bar{X}] &= \frac{\sigma^2}{n}
 \end{aligned}$$

2, A

[1 mark for using property of variance in second line correctly and 1 mark for mentioning that random variables need to be independent for expression in third line to hold.]

sim. seen ↓

(ii)

$$\begin{aligned}
 \text{E}[\bar{X}] &= \text{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\
 &= \frac{1}{n} \sum_{i=1}^n \text{E}[X_i] && \text{(by the linearity of expectation)} \\
 &= \frac{1}{n} \sum_{i=1}^n \mu_i \\
 \Rightarrow \text{E}[\bar{X}] &= \bar{\mu}.
 \end{aligned}$$

1, A

In the notes, we derived

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n(\bar{X})^2 \right),$$

and also the relation, for any random variable Y with mean μ and variance σ^2 , $\text{E}[Y^2] = (\text{E}[Y])^2 + \text{Var}[Y]$.

1, A

Then

$$\begin{aligned}
E[S^2] &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n(\bar{X})^2\right)\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^n X_i^2 - n(\bar{X})^2\right] \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n E[X_i^2] - nE[(\bar{X})^2]\right) \quad (\text{by linearity of expectation}) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n (\{E[X_i]\}^2 + \text{Var}[X_i]) - n(\{E[\bar{X}]\}^2 + \text{Var}[\bar{X}])\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n (\mu_i^2 + \sigma^2) - n(\{\bar{\mu}\}^2 + \sigma^2/n)\right) \quad (\text{using Part (i)}) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n (\mu_i^2) + n\sigma^2 - n\{\bar{\mu}\}^2 - \sigma^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n (\mu_i^2) - n\{\bar{\mu}\}^2 + (n-1)\sigma^2\right) \\
&= \sigma^2 + \frac{1}{n-1}\left(\sum_{i=1}^n (\mu_i^2) - n\{\bar{\mu}\}^2\right) \\
&= \sigma^2 + \frac{1}{n-1}\sum_{i=1}^n [\mu_i - \bar{\mu}]^2.
\end{aligned}$$

3, A

[1 mark for linearity of expectation step, 1 mark for correct substitution using Part (i), 1 mark for computing expression correctly.]

- (b) (i) Since the variance σ^2 is unknown, but the sample variance s^2 is available, we can use Student's t -distribution to construct a confidence interval for μ :

sim. seen ↓

$$T = \frac{\mu - \bar{X}}{S/\sqrt{n}},$$

where the sample mean $\bar{X}$ and the sample variance S^2 are defined by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

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

For a $(1 - \alpha)$ -confidence interval, we note $t_{n-1, 1-\alpha/2} = -t_{n-1, \alpha/2}$ and then

1, A

$$\begin{aligned} P(t_{n-1, \alpha/2} < T < t_{n-1, 1-\alpha/2}) &= 1 - \alpha \\ P\left(-t_{n-1, 1-\alpha/2} < \frac{\mu - \bar{X}}{S/\sqrt{n}} < t_{n-1, 1-\alpha/2}\right) &= 1 - \alpha \\ P\left(-t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} < \mu - \bar{X} < t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}\right) &= 1 - \alpha \\ P\left(\bar{X} - t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}\right) &= 1 - \alpha. \end{aligned}$$

1, A

Since $1 - \alpha = 0.90$, this implies $\alpha = 0.1 \Rightarrow \alpha/2 = 0.05 \Rightarrow 1 - \alpha/2 = 0.95$, and since $n = 12 \Rightarrow n - 1 = 11$, and so $t_{n-1, 1-\alpha/2} = t_{11, 0.95} = 1.796$ from Table 2.

1, A

Therefore, the realised confidence interval for the mean μ is

$$\begin{aligned} \left(\bar{x} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}\right) &= \left(181 - (1.796) \frac{\sqrt{3}}{\sqrt{12}}, 181 + (1.796) \frac{\sqrt{3}}{\sqrt{12}}\right) \\ &= \left(181 - (1.796) \frac{\sqrt{3}}{2\sqrt{3}}, 181 + (1.796) \frac{\sqrt{3}}{2\sqrt{3}}\right) \\ &= (181 - (1.796)/2, 181 + (1.796)/2) \\ &= (181 - 0.898, 181 + 0.898) \\ &\approx (180.102, 181.898), \end{aligned}$$

where any of the numerical expressions on the right hand side of the above equation are acceptable.

1, A

- (ii) There is a theorem in the notes which proves, for $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, that

sim. seen $\Downarrow$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

1, A

We can construct a $(1 - \alpha)$ -confidence interval as follows:

$$\begin{aligned} P\left(\chi_{n-1, \alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1, 1-\alpha/2}^2\right) &= 1 - \alpha \\ P\left(\left\{\chi_{n-1, \alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2}\right\} \text{ and } \left\{\frac{(n-1)S^2}{\sigma^2} < \chi_{n-1, 1-\alpha/2}^2\right\}\right) &= 1 - \alpha \\ P\left(\left\{\frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_{n-1, \alpha/2}^2}\right\} \text{ and } \left\{\frac{1}{\chi_{n-1, 1-\alpha/2}^2} < \frac{\sigma^2}{(n-1)S^2}\right\}\right) &= 1 - \alpha \\ P\left(\frac{1}{\chi_{n-1, 1-\alpha/2}^2} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_{n-1, \alpha/2}^2}\right) &= 1 - \alpha \\ P\left(\frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2}\right) &= 1 - \alpha. \end{aligned}$$

1, B

Since $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$, using Table 3, we compute $n - 1 = 11$ and $\alpha/2 = 0.025$, then $\chi_{n-1, \alpha/2}^2 = \chi_{11, 0.025}^2 = 3.816$, and $1 - \alpha/2 = 0.975$, so from Table 4 $\chi_{n-1, 1-\alpha/2}^2 = \chi_{11, 0.975}^2 = 21.920$.

2, A

Then the realised confidence interval is

$$\begin{aligned} \left(\frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}\right) &= \left(\frac{(11)(3)}{(21.920)}, \frac{(11)(3)}{(2.816)}\right) \\ &\approx (1.505, 11.719), \end{aligned}$$

where the final expression can be left as a pair of fractions.

- (c) The true value of θ may or may not be contained within the computed 99% confidence interval (3, 7).

sim. seen ↓

However, if the experiment could be repeated a large number of times and a new 99% confidence interval be constructed for each repetition of the experiment, then the true value of θ would be in approximately 99% of the constructed confidence intervals.

3, B

[1 mark for noting that θ may or may not be in the confidence interval. 1 mark for mentioning repeated experiments and 1 mark for mentioning that true value of θ would be in 99% of the constructed intervals.]

- (d) The Q-Q plot suggests that the data do follow a normal distribution, since most of the data lie on the line $y = x$, except for the extreme tails.

sim. seen ↓

2, C

[1 mark for correctly interpreting the plot as suggesting the data may follow a normal distribution, and 1 mark for giving reasons.]

5. (a) (i) We first sort the data in non-decreasing order:

sim. seen ↓

$\{0.8, 0.9, 1.1, 1.7, 2.0, 3.0, 3.4, 3.9, 4.5, 4.9, 5.3, 7.5, 9.8\}$.

Since there are $n = 13$ values, the median is the $i_m = (13 + 1)/2 = 7$ th value in this non-decreasing sequence x , i.e. $m = x_{(7)} = 3.4$.

1, A

- (ii) The lower quartile will have index

$i_{0.25} = \frac{1}{2}(\lfloor i_m \rfloor + 1) = \frac{1}{2}(7 + 1) = 4$, and the 4th smallest value in the dataset is $q_{0.25} = x_{(4)} = 1.7$.

1, A

The upper quartile will therefore have index

$i_{0.75} = n - i_{0.25} + 1 = 13 - 4 + 1 = 10$, and so $q_{0.75} = x_{(10)} = 4.9$.

1, A

- (iii) Tukey's criterion defines an outlier to be a value outside the interval

$$[q_{0.25} - 1.5 \cdot \text{IQR}, q_{0.75} + 1.5 \cdot \text{IQR}],$$

where IQR is the interquartile range, i.e. $\text{IQR} = q_{0.75} - q_{0.25} = 4.9 - 1.7 = 3.2$, for this data.

Therefore, $1.5 \cdot (3.2) = 4.8$, and so the interval above becomes

$$[1.7 - 4.8, 4.9 + 4.8] = [-3.1, 9.7].$$

The value 9.8 is outside this interval and is therefore the only outlier.

1, A

- (iv) The boxplot.

seen ↓

1, A

- (b) The null hypothesis H_0 and the alternative hypothesis H_1 for this experiment are

sim. seen ↓

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2.$$

We shall assume that the data $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_m$ are observations of the random variables $X_1, \dots, X_n \sim N(\mu_1, \sigma^2)$ and $Y_1, \dots, Y_m \sim N(\mu_2, \sigma^2)$, and the random variables $X_1, \dots, X_n, Y_1, \dots, Y_m$ are all independent.

1, B

From a result in the notes, $\bar{X} \sim N(\mu_1, \sigma^2/n)$ and $\bar{Y} \sim N(\mu_2, \sigma^2/m)$. Furthermore, $-\bar{Y} \sim N(-\mu_2, \sigma^2/m)$. Moreover, since all the random variables are independent, $\bar{X}$ and $-\bar{Y}$ are also independent, and normal, and so their sum $\bar{X} - \bar{Y}$ is also a normal random variable, by a result in the notes.

1, B

Consider the random variable $\bar{X} - \bar{Y}$. Then using linearity of expectation and a property of the variance,

$$E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_1 - \mu_2,$$

$$\text{Var}[\bar{X} - \bar{Y}] = \text{Var}[\bar{X}] + \text{Var}[\bar{Y}] = \sigma_1^2/n + \sigma_2^2/m.$$

Therefore, $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/n + \sigma_2^2/m)$.

1, B

Under H_0 , $\mu_1 - \mu_2 = 0$, and so we define the random variable Z ,

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}},$$

which from the above discussion implies that $Z \sim N(0, 1)$.

1, B

Given that $\sigma_1^2 = 10$ and $\sigma_2^2 = 4$, $\bar{x} = 185$ and $\bar{y} = 182$ and $n = 15$ and $m = 12$, the realised value of z is

$$\begin{aligned} z &= \frac{185 - 182}{\sqrt{\frac{10}{15} + \frac{4}{12}}} \\ &= \frac{3}{\sqrt{\frac{2}{3} + \frac{1}{3}}} \\ &= \frac{3}{\sqrt{1}} \\ &= 3. \end{aligned}$$

1, A

Using Table 1, we see that $F_Z(3.0) = 0.9987$. Since this is a two-sided test, the p -value is

$$p = 2 \cdot \min\{F_Z(3), 1 - F_Z(3)\} = 2 \cdot \min\{0.9987, 0.0013\} = 2(0.0013) = 0.0026.$$

Because $0.0026 < 0.01$ ($p < \alpha$), the null hypothesis would be rejected.

2, C

[Alternatively, without computing the p -value, 1 mark can be awarded for: For $\alpha = 0.01$, the critical value from Table 1 is $z_{1-\alpha/2} = z_{0.995} = 2.57$ (or some value between 2.57 and 2.58). Since $|3| > 2.57$, we reject the null hypothesis at significance level $\alpha = 0.01$.]

- (c) A Type I error occurs when the null hypothesis has been rejected, when it is in fact true.

seen ↓

1, B

A Type II error occurs when the null hypothesis fails to be rejected, when it is in fact false.

1, B

If a significance threshold α is used, then the probability of a Type I error occurring is $P(\text{Reject } H_0 \mid H_0 \text{ is true}) \leq \alpha$. If α increases, then the probability of a Type I error occurring will necessarily increase.

1, B

Suppose we denote the probability of a Type II error occurring as $P(\text{Fail to reject } H_0 \mid H_0 \text{ is false}) = \beta$. Although α is not directly related to β , if the significance threshold α increases, it will be easier to reject the null hypothesis and so the probability of a Type II error occurring, i.e. β , will decrease.

1, D

sim. seen ↓

- (d) Although four of the five p -values in the table are below the threshold $\alpha = 0.05$, we need to account for and include a correction for multiple hypothesis testing.

1, D

Since there were 10 separate experiments, if we use the Bonferroni correction then we need to use the adjusted significance threshold $\alpha' = \alpha/10 = 0.05/10 = 0.005$.

2, B

Comparing the p -values to the adjusted significance threshold $\alpha' = 0.005$, we see that only $0.002 < \alpha'$, and so only that p -value would be declared to be significant.

1, A

6. (a) Suppose that $\mathbf{x} = (x_1, x_2, \dots, x_n)$ are observations of the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$. The statistician (in the question) assumes that these are independent random variables distributed according to a uniform distribution on the interval $[\theta_1, \theta_2]$, where $\theta_1, \theta_2 \in (0, \infty)$ are unknown parameter values.

unseen ↓

The probability density function of each X_i is

$$f(x_i|\theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \text{for } \theta_1 \leq x_i \leq \theta_2, \\ 0, & \text{otherwise.} \end{cases}$$

1, A

Therefore, the joint likelihood is

$$\begin{aligned} L(\theta_1, \theta_2|\mathbf{x}) &= f(\mathbf{x}|\theta_1, \theta_2) \\ &= \prod_{i=1}^n f(x_i|\theta_1, \theta_2) \quad (\text{by independence}) \\ &= \prod_{i=1}^n \frac{1}{\theta_2 - \theta_1} \\ &= \frac{1}{(\theta_2 - \theta_1)^n}, \end{aligned}$$

when $\theta_1 \leq x_i \leq \theta_2$ for all $i \in \{1, 2, \dots, n\}$.

1, D

This is equivalent to $\theta_1 \leq \min\{x_1, \dots, x_n\} < \max\{x_1, \dots, x_n\} \leq \theta_2$.

1, D

The likelihood $L(\theta_1, \theta_2|\mathbf{x})$ will therefore be maximised when the difference $\theta_2 - \theta_1$ is minimised. This will happen when θ_2 is as small as possible and when θ_1 is as large as possible, which occurs when $\hat{\theta}_2 = \max\{x_1, \dots, x_n\}$ and $\hat{\theta}_1 = \min\{x_1, \dots, x_n\}$.

2, D

The associated maximum likelihood estimators are therefore

$$\hat{\theta}_1 = \min\{X_1, \dots, X_n\}$$

and

$$\hat{\theta}_2 = \max\{X_1, \dots, X_n\}.$$

1, B

(b) (i) Because

seen $\Downarrow$

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0,$$

one can rewrite S_{xy} as

$$\begin{aligned} S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum_{i=1}^n (x_i - \bar{x})y_i - \bar{y} \cdot \sum_{i=1}^n (x_i - \bar{x}) \\ &= \sum_{i=1}^n (x_i - \bar{x})y_i - \bar{y} \cdot 0 \\ &= \sum_{i=1}^n (x_i - \bar{x})y_i. \end{aligned}$$

1, A

(ii) Since

sim. seen $\Downarrow$

$$\begin{aligned} \hat{\beta}_1 &= \hat{\beta}_1(\mathbf{x}, \mathbf{y}) = \frac{S_{xy}}{S_{xx}} \\ &= \frac{1}{S_{xx}} [S_{xy}] \\ &= \frac{1}{S_{xx}} \left[\sum_{i=1}^n (x_i - \bar{x})y_i \right] \quad (\text{from Part (i)}) \\ \Rightarrow \hat{\beta}_1(\mathbf{x}, \mathbf{y}) &= \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})y_i \\ \Rightarrow \hat{\beta}_1 &= \hat{\beta}_1(\mathbf{x}, \mathbf{Y}) = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})Y_i = \sum_{i=1}^n (x_i - \bar{x})Y_i / S_{xx}. \end{aligned}$$

1, A

(iii) Using the expression in (i),

sim. seen ↓

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \left(\frac{S_{xy}}{S_{xx}} \right) \bar{x} \\ &= \bar{y} - \frac{1}{S_{xx}} \left(\sum_{i=1}^n (x_i - \bar{x}) y_i \right) \bar{x} \\ &= \sum_{i=1}^n \frac{y_i}{n} - \frac{1}{S_{xx}} \left(\sum_{i=1}^n (x_i - \bar{x}) \bar{x} \right) y_i \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{S_{xx}} (x_i - \bar{x}) \bar{x} \right) y_i,\end{aligned}$$

and so one can conclude that the maximum likelihood estimator is

$$\hat{\beta}_0 = \sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{S_{xx}} (x_i - \bar{x}) \bar{x} \right) Y_i.$$

1, C

It was given in Part (ii) that the maximum likelihood estimator of $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \frac{1}{S_{xx}} \sum_{j=1}^n (x_j - \bar{x}) Y_j.$$

We also recall (or can deduce) that the $Y_1, Y_2, \dots, Y_n$ are independent and identically distributed according to a $N(\beta_0 + \beta_1 x_i, \sigma^2)$ distribution, so that

$$\text{Cov}(Y_i, Y_j) = \begin{cases} \text{Var}(Y_i) = \sigma^2, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

1, B

Therefore, the covariance of the two estimators can be computed using the bilinearity of the covariance operator as

$$\begin{aligned}\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \text{Cov} \left(\sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{S_{xx}} (x_i - \bar{x}) \bar{x} \right) Y_i, \frac{1}{S_{xx}} \sum_{j=1}^n (x_j - \bar{x}) Y_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{n} - \frac{1}{S_{xx}} (x_i - \bar{x}) \bar{x} \right) \frac{1}{S_{xx}} (x_j - \bar{x}) \text{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{S_{xx}} (x_i - \bar{x}) \bar{x} \right) \frac{1}{S_{xx}} (x_i - \bar{x}) \text{Cov}(Y_i, Y_i) \\ &= \sum_{i=1}^n \left(\frac{1}{n} \frac{1}{S_{xx}} (x_i - \bar{x}) - \frac{1}{(S_{xx})^2} (x_i - \bar{x})^2 \bar{x} \right) \text{Cov}(Y_i, Y_i) \\ &= \left(\frac{1}{n} \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) - \frac{\bar{x}}{(S_{xx})^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right) \sigma^2 \\ &= \left(\frac{1}{n} \frac{1}{S_{xx}} \cdot 0 - \frac{\bar{x}}{(S_{xx})^2} S_{xx} \right) \sigma^2 \\ \Rightarrow \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \frac{-\bar{x} \sigma^2}{S_{xx}}.\end{aligned}$$

2, C

- (iv) One real-world example of linear regression that was covered in the notes was to investigate the (linear) relationship between altitude above sea level and the boiling point of water. Another possible real-world example is investigating the relationship between shoe size and height of a person.

sim. seen ↓

2, C

If there is no clear linear relationship between X and Y , one could use linear regression to see if there is a relationship between $g(X)$ and $h(Y)$, where g and h are transformations such as $g(x) = x^2$ or $h(y) = \log(y)$.

2, D

unseen ↓

- (c) Let the test scores from the first batch of marked scripts be $x_1, x_2, \dots, x_{25}$. One computes bootstrap samples of the data,

$$(x_1^{*j}, x_2^{*j}, \dots, x_{25}^{*j}), \quad j = 1, 2, \dots, B,$$

for some large value B . For example, we could use $B = 1000$.

1, D

Each bootstrap sample has the same number of observations (i.e. 25) as the original sample, and each value in the bootstrap sample is a value from the original sample, i.e. the original data is sampled with replacement, using a discrete uniform distribution on the values $x_1, x_2, \dots, x_{25}$.

1, D

For the bootstrap sample j , we compute the median m^{*j} , so that we have B (bootstrap) medians $m^{*1}, m^{*2}, \dots, m^{*B}$.

1, D

We then use the B bootstrap medians to construct a 95% confidence interval, by sorting them as $m^{*(1)} \leq m^{*(2)} \leq \dots \leq m^{*(B)}$ and setting the confidence interval to be

$$\left(m^{*(\lfloor B \cdot (0.025) \rfloor)}, m^{*(\lfloor B \cdot (0.975) \rfloor)} \right).$$

1, D

Review of mark distribution:

Total A marks: 48 of 48 marks

Total B marks: 30 of 30 marks

Total C marks: 18 of 18 marks

Total D marks: 24 of 24 marks

Total marks: 120 of 120 marks

MATH40005 Probability and Statistics Markers comments

- Question 1
- (a)(i): Most students clearly explained the sample space and probability measure. However, some did not specify what the event space is for this particular experiment, and instead gave only a generic definition. A complete answer required identifying the event space relevant to the die-rolling scenario.
- (a)(ii), parts (a) and (b): These parts were generally answered well. A common error was that some students assumed a six-sided die instead of the n -sided die specified in the question.
- (b): This question was also handled well overall. However, some students incorrectly focused on the lateness of train services rather than the probability that a passenger arrives late, which the question explicitly asked for. Additionally, a few students did not specify which theorem (e.g. law of total probability or Bayes' theorem) they were applying.
- Question 2
- 2(a) This was a classic combinatorics question involving constrained arrangements with character types. A common mistake was assuming the characters were unordered, which is not the case when dealing with passwords.
- 2(b) This part was overall done well; students were familiar with similar problems. Some common mistakes included failing to check non-negativity condition (part (i)) and not clearly stating the ranges of the random variables involved (parts (ii) and (iii)). Part (iv) even though this question did not require students to compute probabilities numerically or derive marginal densities to determine independence, they were not penalised for including such calculations.
- Question 3
- Part (a) was answered well overall, with students mainly losing marks due to not justifying their steps. Similarly for part (b), where the MGF of the Normal was proven to be trickier than expected and many students did not use that to answer (b)(ii), despite being asked to do so. Most students had the right idea of what to write in part (c)(i) but were not very clear in their proofs, while (c)(ii) was not answered well, mainly due to students either skipping it or not providing a reasoning.

- Question 4
- 4(a)(i) was generally answered well, as expected since it was book work, although often independence was not mentioned as justification when using the property that the variance of a sum is the sum of the variances.
- 4(a)(ii) was generally not answered well, even though it is a small variation of a result and proof in the notes.
- 4(b)(i) and Q4(b)(ii) were generally answered very well.
- 4(c) was generally answered well, although there were many cases of students saying that "99% of the constructed parameters θ are in (3, 7)", rather than the correct "99% of the constructed intervals contain the true value of θ ".
- 4(d) was generally not answered well, with many students saying that the few points at the tails did not lie on the line $y=x$ meant the data was probably *not* normally distributed.
- Question 5
- All parts of Q5(a) were typically well answered, though some students incorrectly used $\text{median} \pm 1.5 \times \text{IQR}$ to identify outliers, and there were occasional arithmetic errors.
- Part (b) was more challenging, with some students struggling to identify the correct distribution and test statistic for hypothesis testing—common errors included using variants of the Student test or chi-squared tests. Many students also opted to compare the test statistic to a critical value rather than computing a p-value. For part (c), most students correctly defined Type I and Type II errors, though a common mistake is the link between the significance threshold with the probability of a Type II error.
- Part (d) was mostly well done, though some answers lacked justification for the correction method used.
- 6a) Students were generally able to compute the likelihood, but often struggled to compute the maximum likelihood.
- 6b) Parts (i) and (ii) were done well, but students often struggled to compute the covariance in part (iii); in part (iv) students were generally able to provide good real-world examples of linear regression, but often did not mention the possibility of transforming the variables.
- Question 6
- 6c) Students generally knew how to compute bootstrap samples but some struggled to explain how to use these samples to construct a confidence interval.