

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2021

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

Probability and Statistics

Date: Thursday, 13 May 2021

Time: 09:00 to 12:00

Time Allowed: 3 hours

Upload Time Allowed: 45 minutes

This paper has 6 Questions.

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

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS SEPARATE PDFs TO THE RELEVANT DROPBOXES ON BLACKBOARD (ONE FOR EACH QUESTION) WITH COMPLETED COVERSHEETS WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

Throughout the exam, we assume that $(\Omega, \mathcal{F}, P)$ denotes a probability space.

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

1. (a) Consider a sample space given by $\Omega = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Consider the set $A = \{(x, y) \in \Omega : y \leq 0.5x\}$. Sketch Ω and A and explain how you could assign a probability to the event A . (2 marks)
- (b) Consider the experiment where you roll a six-sided fair die seven times.
 - (i) Find the sample space Ω of this experiment. (2 marks)
 - (ii) What is the cardinality of the power set of Ω ? (2 marks)
 - (iii) Find a probability space for this experiment. (2 marks)
- (c) Consider independent events $A, B \in \mathcal{F}$. Are their complements A^c and B^c independent? If so, prove it, otherwise give a counterexample. (2 marks)
- (d) The English alphabet has 26 letters (5 vowels and 21 consonants).
 - (i) How many 7 letter words are there with 5 distinct consonants and 2 distinct vowels? (3 marks)
 - (ii) How many 7 letter words are there with 5 distinct consonants and 2 (not necessarily distinct) vowels? (3 marks)

In both (i) and (ii) the order of the letters matters and "words" include imaginary words.
- (e) Ten cars race in a car race under ideal conditions. The only difference between the cars is their colour: Three cars are red, one car is white, two cars are yellow, and four cars are green. What is the cardinality of the sample space of all the possible race outcomes (assuming cars of the same colour are indistinguishable)? (2 marks)
- (f) A boy needs to wear his school uniform (pants, shirt, tie and jacket) to school. He owns 3 pairs of pants, 4 shirts, 5 ties and 3 jackets. How many different outfits can he wear to school if he must wear one of each item?

(2 marks)

(Total: 20 marks)

2. (a) Suppose one fair six-sided die is rolled twice. Let X_1 and X_2 denote the score of the die in the first and the second throw, respectively. For $i, j \in \{1, 2, 3, 4, 5, 6\}$, find

$$P(\max(X_1, X_2) = i | X_1 = j).$$

(3 marks)

- (b) Let X and Y be independent, discrete random variables taking values in $\mathbb{N} = \{1, 2, \dots\}$ with identical probability mass functions, for $p \in (0, 1)$, given by

$$p_X(z) = P(X = z) = (1 - p)^{z-1}p,$$

$$p_Y(z) = P(Y = z) = (1 - p)^{z-1}p,$$

for $z \in \mathbb{N}$, $p_X(z) = p_Y(z) = 0$ otherwise. Find

$$E\left(\frac{X^2 + Y^2}{XY}\right).$$

(6 marks)

- (c) Let (X, Y) be two jointly continuous random variables with joint probability density function $f_{X,Y}$ and marginal probability densities denoted by f_X and f_Y , respectively. Let $F_{Y|X=x}$ denote the conditional distribution function of Y given $X = x$. Show that

$$P(Y \leq y) = \int_{\{x: f_X(x) > 0\}} F_{Y|X=x}(y|x) f_X(x) dx, \quad y \in \mathbb{R}.$$

Hint: Here you may interchange the order of integration in any multiple integrals without justification.

(4 marks)

- (d) Let X be a standard normal random variable with moment generating function $M_X(u) = E(e^{uX}) = \exp(u^2/2)$, for $u \in \mathbb{R}$, and let Y be a random variable with Gamma(α, β) distribution, i.e. its probability density function is given by $f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$ for $y > 0$ and 0 otherwise. Suppose that X and Y are independent. Define the continuous random variable $Z = \sqrt{Y}X$. Find the moment generating function of Z .

(7 marks)

(Total: 20 marks)

3. (a) Let X denote a continuous random variable with uniform distribution on the interval (a,b) for $a < b$, i.e. its density function is given by $f_X(x) = (b - a)^{-1}$ for $a < x < b$ and 0 otherwise. Define $Y = X - 1$.
- (i) Find the cumulative distribution function of Y . (2 marks)
 - (ii) Is Y a continuous random variable? Justify your answer. (3 marks)
 - (iii) Are the mean and median of Y identical? Justify your answer. (4 marks)
- (b) Explain when you would typically apply the law of the unconscious statistician and describe briefly why it is useful. (3 marks)
- (c) Consider two jointly continuous random variables X, Y with joint probability density function given by

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

- (i) For which choice of the constant c is $f_{X,Y}$ a valid density? (1 mark)
- (ii) Find the marginal densities of X and Y . (3 marks)
- (iii) Find

$$P\left(\frac{X}{Y} \leq 0.5\right).$$

(4 marks)

(Total: 20 marks)

4. (a) Suppose that $X_1, X_2, \dots, X_n$ are independent random variables which all follow the same distribution with unknown mean μ and unknown variance σ^2 . Suppose that X is a random variable that also follows this distribution and let $\mu'_2 = E[X^2]$ denote the second moment of this distribution.
- (i) Stating any identities used, show that $\mu'_2 = \mu^2 + \sigma^2$. (1 mark)
 - (ii) Show that $\Delta = \frac{1}{n} \sum_{i=1}^n X_i^2$ is an unbiased estimator of μ'_2 . (2 marks)
 - (iii) Define an estimator $\hat{\Theta} = \hat{\Theta}(X_1, X_2, \dots, X_n)$ of $(E[X])^2$ in terms of the random sample $X_1, X_2, \dots, X_n$ **such that $\hat{\Theta}$ is an unbiased estimator**, and moreover, $\hat{\Theta}$ is not constant in any of the variables. (For example, $\hat{\Theta} = X_1^2$ is constant in all of the variables except X_1 , so this would not be acceptable). (3 marks)
 - (iv) Suppose that $n = 2$ and we observe (X_1, X_2) as (x_1, x_2) . Using your estimator in Part (iii), compute $\hat{\theta}(x_1, x_2)$, the realisation of your estimator $\hat{\Theta}(X_1, X_2)$, when observing $x_1 = 3$ and $x_2 = -1$. Is there any problem with your estimator? (2 marks)
- (b) A farmer harvests his crop of over one million potatoes, and wishes to know how much each potato weighs on average. He would be satisfied by knowing the average weight of each potato to within a few grams with high confidence. From previous years, he is certain that the standard deviation in weight of the potatoes is less than 10 grams. What is the minimum number of randomly selected potatoes he would need to weigh in order to be 99% confident that he will estimate the average weight of a single potato to within 2 grams, i.e. that the confidence interval he computes will be less than 4 units wide (measured in grams)? (6 marks)
- (c) Suppose that the following data are recorded:

{47.5, 50.7, 46.7, 56.4, 51.3, 46.7, 62.4, 53.0, 52.3}.

Showing all working:

- (i) Compute the median of this dataset. (1 mark)
- (ii) Compute the lower quartile $q_{0.25}$ and the upper quartile $q_{0.75}$ of this dataset. (2 marks)
- (iii) One popular rule for determining if any observation x in a dataset is an outlier is the following:

$$x \notin [q_{0.25} - 1.5 \cdot \text{IQR}, q_{0.75} + 1.5 \cdot \text{IQR}] \Rightarrow x \text{ is an outlier}$$

where IQR is the interquartile range of the dataset. Using this rule, determine whether or not any of the observations in this dataset are outliers. (2 marks)

- (iv) Which visualisation would be suitable for graphically displaying the statistics and outliers (if any) computed in Parts (i)-(iii)? (1 mark)

(Total: 20 marks)

5. (a) Suppose that the random variables $X_1, X_2, \dots, X_n$ are independent and identically distributed according to a probability distribution with probability density function given by

$$f_X(x) = \begin{cases} \theta x^{\theta-1}, & \text{for } 0 \leq x \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is an unknown parameter. Showing full working, find the maximum likelihood estimate of θ . (6 marks)

- (b) In a large class of students, the heights of the students are assumed to follow a normal distribution with unknown mean μ and unknown variance σ^2 . The students guess that the average height of a student in the class is 175cm. They measure the heights of 18 randomly-selected students, where these heights are labelled $x_1, x_2, \dots, x_{18}$, and the sample mean of these measurements is computed to be $\bar{x} = 176.8\text{cm}$ and the sample variance is computed to be $s^2 = 8$.
- (i) In order to test the students' hypothesis that the average height of a student in the class is 175cm, write down appropriate null and alternative hypothesis. (1 mark)
 - (ii) Using a significance threshold of $\alpha = 0.01$, compute the appropriate test statistic and critical threshold for the test statistic, based on your hypotheses in Part (i). Note that Tables 1 and 2 may be helpful, and you may use a calculator. (3 marks)
 - (iii) Based on the values computed in Part (ii), make a careful statement about whether or not the average student height is 175cm. (2 marks)
 - (iv) Again using $\alpha = 0.01$, construct a $1 - \alpha$ confidence interval for the average height of a student in the class. You may leave the expression unsimplified, or you may use a calculator to compute the limits to two decimal places. (1 mark)

Table 1: Partial table showing values of t for $P(T < t)$, where T has Student's t -distribution with ν degrees of freedom

ν	0.95	0.975	0.99	0.995
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845

Table 2: Partial table showing values of z for $P(Z < z)$, where Z has a standard normal distribution

z	$P(Z < z)$
1.281	0.900
1.645	0.950
1.960	0.975
2.326	0.990
2.576	0.995

Note: Question 5 continues on the next page.

- (c) Briefly describe a real-world example where Student's two-sample test could be used, making reference to any assumptions that the test requires. (3 marks)
- (d) Suppose a pharmaceutical company runs 200 clinical trials, where each clinical trial tests a different drug for curing a particular disease, and the null hypothesis for each trial is that the drug does not cure the disease. For each clinical trial, a statistical analysis is performed on the collected data which results in a p -value being calculated. Assuming that none of the trialled drugs actually cure the disease (i.e. the null hypothesis is true for all trials), how many of the 200 p -values should we expect to be below the threshold $\alpha = 0.05$? Provide full justification for your answer. (4 marks)

(Total: 20 marks)

6. (a) Recall that the simple linear regression model for data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ is given by

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where the $\epsilon_i \sim N(0, \sigma^2)$ are independent. Recall also that the maximum likelihood estimates for the parameters β_0 and β_1 are the constants

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

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ and

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2.$$

- (i) Show that $\text{Var}(Y_i) = \sigma^2$. (1 mark)
(ii) Show that S_{xy} can be rewritten as

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

and so the estimator for $\hat{\beta}_1$ can be rewritten as $\hat{\beta}_1 = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})Y_i$. (2 marks)

- (iii) Showing full working, compute

$$\text{Cov}(Y_i, \hat{\beta}_1),$$

for $i = 1, 2, \dots, n$. (4 marks)

- (iv) Briefly describe a real-world example where linear regression can be used, making reference to how the estimates of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ can be used. (3 marks)

- (b) Suppose that the random variables $X_1, X_2, \dots, X_n$ are known to follow a distribution with probability mass function

$$f(x|\theta) = \theta(1 - \theta)^{x-1}, \quad \text{where } x = 1, 2, 3, \dots,$$

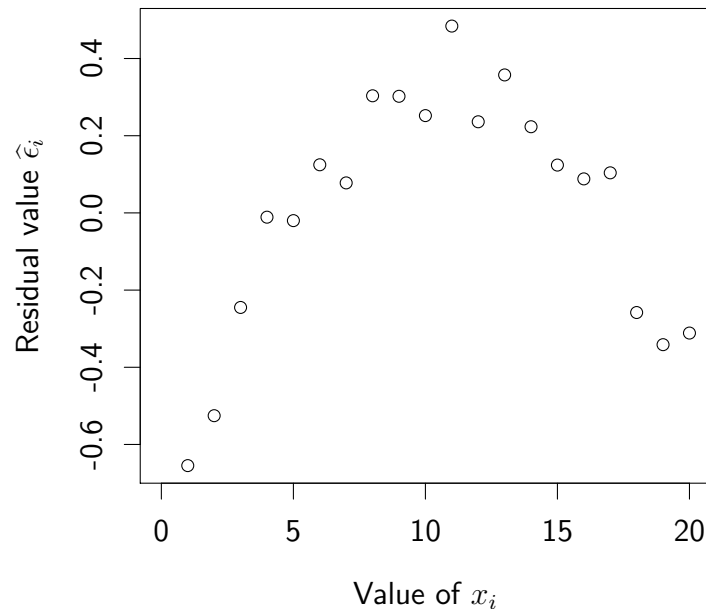
where $\theta \in (0, 1)$. Following a Bayesian approach, assume that the prior distribution of θ follows a distribution with probability density function

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \quad \text{where } \theta \in (0, 1).$$

Given that the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$ are observed as $\mathbf{x} = (x_1, x_2, \dots, x_n)$, find $\pi(\theta|\mathbf{x})$, the probability density function of the posterior distribution of θ given the data $\mathbf{x}$, and identify this distribution (state the family of distributions to which it belongs as well as the value of its parameters).

(5 marks)

- (c) Suppose one fits a simple linear regression model to the data $(x_1, y_1), (x_2, y_2), \dots, (x_{20}, y_{20})$, and obtains estimates for the regression coefficients $\hat{\beta}_0 = 9.445$ and $\hat{\beta}_1 = 0.379$, and the value of R^2 is $R^2 = 0.984$. Given this information and the residual plot below, briefly assert and justify whether or not the model fits the data well. (2 marks)



- (d) A new drug is developed to prevent the development of a disease. A clinical trial is run, where some patients receive a placebo (a pill that has no medical effect), while others receive the new drug. The trial was split into two groups of patients, and Table 3 below shows the results. When the groups are combined, the data shows that patients taking the placebo have a lower chance of developing the disease. However, within each each group, patients who took the drug have a lower chance of developing the disease.

	Placebo (no drug)	Drug
Group A	5/10=0.5	14/50=0.28
Group B	7/50=0.14	1/10=0.1
Combined	12/60=0.2	15/60=0.25

Table 3: Proportion of patients who develop the disease.

- What is the name of this statistical phenomenon? (1 mark)
- What conditions cause this phenomenon to occur? (1 mark)
- In this example, describe how the groups could have been chosen so that, despite the combined results appearing to show that the placebo is better, it is actually more beneficial for a patient to take the drug. (1 mark)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2021

This paper is also taken for the relevant examination for the Associateship.

M40005

Probability and Statistics (Solutions)

Setter's signature

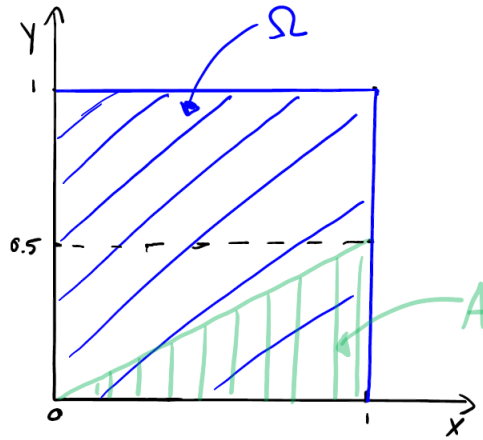
.....

Checker's signature

.....

Editor's signature

.....



meth seen ↓

1, A

1. (a)

We note that the area of Ω is finite and equal to 1. The event A is a subset of Ω . We could assume that the probability of the event A should be uniform on Ω , meaning that we could assign a probability, using the classical interpretation, as

$$P(A) = \frac{\text{area of } A}{\text{area of } \Omega} = \frac{1}{4}.$$

1, A

[Note that there are many possible alternative solutions. E.g. one could assign an arbitrary number between 0 and 1 to $P(A)$ if one, say, considers a σ -algebra $\{\emptyset, A, A^c, \Omega\}$ as the corresponding event space.]

meth seen ↓

(b) (i) We can express the sample space as a Cartesian product space:

$$\Omega = \underbrace{\{1, 2, 3, 4, 5, 6\} \times \cdots \times \{1, 2, 3, 4, 5, 6\}}_{7 \text{ times}} = \{1, 2, 3, 4, 5, 6\}^7.$$

Alternatively, we could write

$$\Omega = \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7) : \omega_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, 2, 3, 4, 5, 6, 7\}.$$

2, A

[Only one answer is required for full marks.]

(ii) By the multiplication principle, we have $\text{card}(\Omega) = 6^7 (= 279936)$.

1, A

From the lecture notes we know that the cardinality of the powerset of Ω is given by $2^{\text{card}(\Omega)} = 2^{(6^7)}$.

1, A

(iii) We can define the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \{1, 2, 3, 4, 5, 6\}^7$, $\mathcal{F} = \mathcal{P}(\Omega)$ (the power set of Ω) and, using the naive/classical interpretation of probability,

1, B

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{6^7},$$

for all $A \in \mathcal{F}$. We know from lectures that the power set is a sigma algebra, and the naive/classical probability measure is a probability measure.

1, B

- (c) Yes, A^c and B^c are independent (as stated in the lecture notes). We need to show that $P(A^c \cap B^c) = P(A^c)P(B^c)$. Since A and B are independent, we know that

meth seen ↓

$$P(A \cap B) = P(A)P(B).$$

Hence, using properties of the probability measure,

$$P(A^c)P(B^c) = (1 - P(A))(1 - P(B)) = 1 - P(A) - P(B) + P(A)P(B) \\ \stackrel{A, B \text{ indep.}}{=} 1 - P(A) - P(B) + P(A \cap B).$$

Also, by De Morgan's law and a result from lectures, we have that

$$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)].$$

Hence

$$P(A^c)P(B^c) = P(A^c \cap B^c).$$

- (d) (i) There are $\binom{21}{5}$ ways of choosing the 5 distinct consonants and $\binom{5}{2}$ ways of choosing two distinct vowels. We can then arrange these seven letters in $7!$ possible ways. So, the final answer is

2, A

unseen ↓

$$7! \binom{21}{5} \binom{5}{2}.$$

- (ii) Let us consider the case when we choose the same vowel twice: There are 5 vowels and, if we choose the same twice, we need to account for over-counting since we cannot distinguish them. So, there are

3, B

$$7! \binom{21}{5} \frac{5}{2!}$$

ways of choosing a 7-letter word with two identical vowels and 5 distinct consonants. Adding this result to the result in (ii) leads to the final answer:

2, C

$$7! \binom{21}{5} \binom{5}{2} + 7! \binom{21}{5} \frac{5}{2!} = 7! \binom{21}{5} \left[\binom{5}{2} + \frac{5}{2} \right] = 7! \binom{21}{5} \cdot 12.5.$$

1, B

- (e) Let Ω be the sample space of all possible race outcomes, where cars of the same colour are indistinguishable. There are three red, one white, two yellow and four green cars. There are $10!$ possible permutations of these 10 cars, and accounting for over-counting, since some have the same colour, we get

$$\text{card}(\Omega) = \frac{10!}{3!1!2!4!} (= 12600).$$

2, B

- (f) The boy needs to choose 1 of each of the four items, i.e. there are

$$\binom{3}{1} \binom{4}{1} \binom{5}{1} \binom{3}{1} = 3 \cdot 4 \cdot 5 \cdot 3 = 180$$

possible outfits.

2, B

2. (a) Let $i, j \in \{1, 2, 3, 4, 5, 6\}$. We consider three cases:

unseen ↓

Case 1: If $i < j$, then

$$P(\max(X_1, X_2) = i | X_1 = j) = 0.$$

Case 2: If $i = j$, then

$$\begin{aligned} P(\max(X_1, X_2) = i | X_1 = j) &= P(\max(X_1, X_2) = j | X_1 = j) \\ &= P(\max(j, X_2) = j | X_1 = j) \stackrel{X_1, X_2 \text{ indep.}}{=} P(\max(j, X_2) = j) = P(X_2 \leq j) = \frac{j}{6}. \end{aligned}$$

Case 3: If $i > j$, then

$$\begin{aligned} P(\max(X_1, X_2) = i | X_1 = j) &= P(\max(j, X_2) = i | X_1 = j) \\ &\stackrel{X_1, X_2 \text{ indep.}}{=} P(\max(j, X_2) = i) = P(X_2 = i) = \frac{1}{6}. \end{aligned}$$

- (b) Since X and Y are independent, their joint probability mass function is given by

3, A

meth seen ↓

$$p_{X,Y}(x, y) = (1-p)^{x-1}p(1-p)^{y-1}p,$$

for $x, y \in \mathbb{N}$ and $p_{X,Y}(x, y) = 0$ otherwise.

Using the law of the unconscious statistician, we get

$$\begin{aligned} A &:= E\left(\frac{X^2 + Y^2}{XY}\right) = \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \frac{x^2 + y^2}{xy} (1-p)^{x-1}p(1-p)^{y-1}p \\ &= \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \frac{x}{y} (1-p)^{x-1}p(1-p)^{y-1}p + \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \frac{y}{x} (1-p)^{x-1}p(1-p)^{y-1}p \\ &= p^2 \sum_{x=1}^{\infty} x(1-p)^{x-1} \sum_{y=1}^{\infty} \frac{1}{y} (1-p)^{y-1} + p^2 \sum_{x=1}^{\infty} \frac{1}{x} (1-p)^{x-1} \sum_{y=1}^{\infty} y(1-p)^{y-1} \\ &= 2p^2 \sum_{x=1}^{\infty} x(1-p)^{x-1} \sum_{y=1}^{\infty} \frac{1}{y} (1-p)^{y-1}. \end{aligned}$$

We use the geometric series and the series expansion of the logarithm to evaluate these sums: For $p \in (0, 1)$, we know that

3, C

$$f(p) := \sum_{x=0}^{\infty} (1-p)^x = p^{-1}, \quad f'(p) = -\sum_{x=1}^{\infty} x(1-p)^{x-1} = -p^{-2}.$$

Hence $\sum_{x=1}^{\infty} x(1-p)^{x-1} = p^{-2}$. Also, using the series expansion of the logarithm, we note that

$$\sum_{x=1}^{\infty} \frac{1}{x} (1-p)^{x-1} = \frac{\log(p)}{p-1}.$$

Altogether, we get

$$A = 2p^2 \frac{\log(p)}{p-1} p^{-2} = \frac{2 \log(p)}{p-1}.$$

3, D

(c) From lectures, we know that, for $y \in \mathbb{R}$,

unseen ↓

$$\begin{aligned} P(Y \leq y) &= \int_{-\infty}^y f_Y(v) dv, \\ f_Y(v) &= \int_{-\infty}^{\infty} f_{X,Y}(x, v) dx, \\ f_{Y|X=x}(y|x) &= \frac{f_{X,Y}(x, v)}{f_X(x)}, \text{ for } x \text{ such that } f_X(x) > 0. \end{aligned}$$

Hence, we can write

$$\begin{aligned} P(Y \leq y) &= \int_{-\infty}^y f_Y(v) dv = \int_{-\infty}^y \int_{-\infty}^{\infty} f_{X,Y}(x, v) dx dv \\ &= \int_{-\infty}^y \int_{\{x: f_X(x) > 0\}} \frac{f_{X,Y}(x, v)}{f_X(x)} f_X(x) dx dv \\ &= \int_{-\infty}^y \int_{\{x: f_X(x) > 0\}} f_{Y|X=x}(v|x) f_X(x) dx dv \\ &= \int_{\{x: f_X(x) > 0\}} \int_{-\infty}^y f_{Y|X=x}(v|x) dv f_X(x) dx \\ &= \int_{\{x: f_X(x) > 0\}} F_{Y|X=x}(y|x) f_X(x) dx. \end{aligned}$$

[The interchange of the integrals can be justified by Tonelli's theorem, this was not covered and hence is not expected for full marks.]

4, D

(d) We use the law of total expectation to derive the result, i.e. we compute

unseen ↓

$$E(e^{uZ}) = \int_0^{\infty} E(e^{uZ} | Y = y) f_Y(y) dy,$$

for u such that the expectation exists.

1, D

Let $u \in \mathbb{R}, y > 0$, then

$$\begin{aligned} E(e^{uZ} | Y = y) &= E(e^{u\sqrt{Y}X} | Y = y) = E(e^{u\sqrt{y}X} | Y = y) \stackrel{X,Y \text{ indep.}}{=} E(e^{u\sqrt{y}X}) \\ &= M_X(\sqrt{y}u) = \exp(u^2 y / 2). \end{aligned}$$

2, D

Let $u \in \mathbb{R}$ (we will later need to restrict this further), by the law of total expectation, we have

$$\begin{aligned} E(e^{uZ}) &= \int_0^{\infty} E(e^{uZ} | Y = y) f_Y(y) dy = \int_0^{\infty} \exp(u^2 y / 2) f_Y(y) dy \\ &= \int_0^{\infty} \exp\left(\frac{u^2}{2} y\right) f_Y(y) dy = M_Y(u^2 / 2), \end{aligned}$$

for $u^2 / 2 < \beta$. Note that the restriction on u comes from the fact that $M_Y(t)$ is defined for $t < \beta$ as the following computation will show.

Also, for $\beta > t$,

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_0^\infty e^{ty} f_Y(y) dy = \int_0^\infty e^{ty} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-(\beta-t)y} dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-(\beta-t)y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\beta-t)^{-\alpha} \int_0^\infty z^{\alpha-1} e^{-z} dz = \frac{\beta^\alpha}{\Gamma(\alpha)} (\beta-t)^{-\alpha} \Gamma(\alpha) = \left(\frac{\beta}{\beta-t} \right)^\alpha. \end{aligned}$$

Hence, for $u^2/2 < \beta$,

$$M_Z(u) = E(e^{uZ}) = \left(\frac{\beta}{\beta - u^2/2} \right)^\alpha.$$

4, C

3. (a) (i) For $y \in \mathbb{R}$, we have

meth seen ↓

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X - 1 \leq y) = P(X \leq y + 1) = F_X(y + 1) \\ &= \begin{cases} 0, & \text{for } y \leq a - 1, \\ \frac{y+1-a}{b-a}, & \text{for } a - 1 < y < b - 1, \\ 1, & \text{for } y \geq b - 1. \end{cases} \end{aligned}$$

2, A

(ii) Yes, Y is a continuous random variable. To see this, note that F_Y can be differentiated for all $y \in \mathbb{R} \setminus \{a - 1, b - 1\}$, with

$$f_Y(y) = \frac{1}{b-a}, \text{ for } a - 1 < y < b - 1,$$

and $f_Y(y) = 0$ otherwise. Since $f_Y(y) \geq 0$ for all $y \in \mathbb{R}$, $\int_{-\infty}^{\infty} f_Y(y) dy = 1$ and $F_Y(y) = \int_{-\infty}^y f_Y(v) dv$, we conclude that Y is a continuous random variable.

3, A

(iii) The mean of Y is given as

$$\begin{aligned} E(Y) &= E(X) - 1 = \int_a^b \frac{x}{b-a} dx - 1 = \frac{x^2}{2(b-a)} \Big|_a^b - 1 = \frac{b^2 - a^2}{2(b-a)} - 1 \\ &= \frac{(b-a)(b+a)}{2(b-a)} - 1 = \frac{b+a}{2} - 1, \end{aligned}$$

where we used the linearity of the expectation. The median of Y is given by the value y satisfying

2, A

$$F_Y(y) = 0.5 \Leftrightarrow \frac{y+1-a}{b-a} = 0.5 \Leftrightarrow y = \frac{1}{2}(a+b) - 1.$$

Hence, we find that the mean and median of Y are identical.

2, B

(b) Let X be a continuous (discrete) random variable with pdf (pmf) f_X (p_X). Suppose we consider a transformation $Y = g(X)$, such that Y is also continuous (discrete). If we want to compute the expectation of the transformed random variable, the law of the unconscious statistician states that

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

in the continuous case or

$$E(g(X)) = \sum_x g(x) p_X(x),$$

in the discrete case, provided the expectations exist. This result is useful since it means that we do not need to derive the pdf/pmf of the transformed random variable Y , which can sometimes be rather tedious.

3, B

(c) (i) We need that

meth seen ↓

$$\int_0^1 \int_0^1 f_{X,Y}(xy) dx dy = 1,$$

which is equivalent to

$$\begin{aligned} c^{-1} &= \int_0^1 \int_0^1 (x+y) dx dy = \int_0^1 \left(\frac{1}{2}x^2 + xy \Big|_0^1 \right) dy = \int_0^1 \left(\frac{1}{2} + y \right) dy \\ &= \frac{1}{2}y + \frac{1}{2}y^2 \Big|_0^1 = 1, \end{aligned}$$

hence we need that $c = 1$.

1, A

(ii) We have

$$f_X(x) = \int_0^1 (x+y) dy = xy + 0.5y \Big|_0^1 = x + 0.5,$$

for $0 < x \leq 1$ and $f_X(x) = 0$ otherwise. By symmetry, $f_Y(y) = y + 0.5$ for $0 < y \leq 1$ and $f_Y(y) = 0$ otherwise.

3, A

(iii) Let $A := \{(x, y) : x/y < 0.5\} = \{(x, y) : x < 0.5y\}$. From a result from lectures, we get that

unseen ↓

$$\begin{aligned} P(X/Y < 0.5) &= \int \int_A f_{X,Y}(x, y) dx dy = \int_0^1 \int_0^{0.5y} (x+y) dx dy \\ &= \int_0^1 \left(0.5x^2 + xy \Big|_0^{0.5y} \right) dy = \frac{1}{8} \frac{1}{3} y^3 + \frac{1}{2} \frac{1}{3} y^3 \Big|_0^1 = \frac{5}{24} (\approx 0.208). \end{aligned}$$

[2 D marks for method, 2 A marks for calculation]

2, D

2, A

4. (a) (i) In the lecture notes (Spring, Exercise 1.1.5), we have the identity

seen ↓

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

Then

$$\mu'_2 = E[X^2] = \text{Var}(X) + (E[X])^2 = \sigma^2 + \mu^2,$$

as required.

1, A

- (ii) Using the linearity of expectation and Part (i),

seen/sim.seen ↓

$$E[\Delta] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^2\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] = \frac{1}{n} \sum_{i=1}^n (\mu'_2) = \frac{1}{n} \cdot n(\mu'_2) = \mu'_2,$$

as required.

2, A

- (iii) Using the identity $\text{Var}(X) = E[X^2] - (E[X])^2$,

unseen ↓

$$(E[X])^2 = E[X^2] - \text{Var}(X).$$

We have shown in Part (ii) that Δ is an unbiased estimator of $E[X^2]$. We know from Proposition 1.2.6 in the Spring notes that the sample variance S^2 is an unbiased estimator of σ^2 . Therefore,

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

is an unbiased estimator of $(E[X])^2$, since

2, D

$$E[\hat{\Theta}] = E[\Delta - S^2] = E[\Delta] - E[S^2] = \mu'_2 - \sigma^2 = \mu^2 = (E[X])^2.$$

Moreover, it is not constant in any of the variables.

1, A

[Note: $(\bar{X})^2$ would not be a suitable estimator, because it is not unbiased.]

meth seen ↓

- (iv) If $n = 2$ and $x_1 = 3$ and $x_2 = -1$, then

$$\bar{x} = \frac{1}{2}[3 - 1] = 1,$$

$$s^2 = \frac{1}{2-1}[(3-1)^2 + (-1-1)^2] = (2)^2 + (-2)^2 = 8$$

$$\Delta(3, -1) = \frac{1}{2}[3^2 + (-1)^2] = \frac{10}{2} = 5,$$

and so

$$\hat{\theta}(3, -1) = 5 - 8 = -3.$$

1, A

This is a problem because $(E[X])^2 \geq 0$ but our estimate is negative.

unseen ↓

1, C

- (b) Let n be the number of potatoes the farmer will weigh. Let the weight of these n potatoes be the random variables $X_1, X_2, \dots, X_n$. Let's assume these weights are independent and identically distributed according to some distribution F_X , which has mean μ and variance σ^2 . Then the sample mean is

sim. seen ↓

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

and $E[\bar{X}] = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, using Proposition 1.2.6 in the notes.

2, A

Although we are not given the distribution F_X , we can use Chebyshev's inequality on the random variable $\bar{X}$ to obtain

$$\begin{aligned} P\left(|\bar{X} - \mu| \geq k \frac{\sigma}{\sqrt{n}}\right) &\leq \frac{1}{k^2} \\ \Rightarrow P\left(|\bar{X} - \mu| < k \frac{\sigma}{\sqrt{n}}\right) &\geq 1 - \frac{1}{k^2} \end{aligned}$$

where k is some parameter.

1, D

If $k = 10$, then

$$1 - \frac{1}{k^2} = 1 - 0.01 = 0.99,$$

and so the interval

$$\left(\bar{x} - 10 \frac{\sigma}{\sqrt{n}}, \bar{x} + 10 \frac{\sigma}{\sqrt{n}}\right)$$

is a 99% confidence interval for the unknown mean μ .

1, B

The width of this interval is

$$20 \frac{\sigma}{\sqrt{n}} < \frac{200}{\sqrt{n}}$$

since we are given that $\sigma < 10$. Therefore, we need to choose n such that

$$\begin{aligned} \frac{200}{\sqrt{n}} &< 4 \\ \Rightarrow \frac{200}{4} &< \sqrt{n} \\ \Rightarrow 50 &< \sqrt{n} \\ \Rightarrow n &> 2500. \end{aligned}$$

So weighing 2501 potatoes will be enough to achieve this accuracy.

2, D

(c) (i) Sorting the data,

meth seen ↓

$\{46.7, 46.7, 47.5, 50.7, 51.3, 52.3, 53.0, 56.4, 62.4\}.$

Since there are nine values, the median is $m = x_{([9+1]/2)} = x_{(5)} = 51.3.$

1, A

(ii) Since there are nine values, the lower quartile is

$$q_{0.25} = x_{([1+5]/2)} = x_{(3)} = 47.5,$$

and the upper quartile is

$$q_{0.75} = x_{([5+9]/2)} = x_{(7)} = 53.0.$$

2, A

(iii) The interquartile range is $IQR = q_{0.75} - q_{0.25} = 53.0 - 47.5 = 5.5$, therefore the interval is

$$[47.5 - 1.5 \cdot 5.5, 53.0 + 1.5 \cdot 5.5] = [39.25, 61.25]$$

1, A

The only observation that is outside this interval is 62.4, so this is the only outlier according to the rule stated.

1, A

(iv) A boxplot would be suitable since it displays the median, lower quartile, upper quartile and any outliers.

1, A

5. (a) Since the random variables are independent, we can compute the likelihood as

meth seen ↓

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1}.$$

We could compute the derivative directly, but it is better to first compute the log-likelihood:

$$\begin{aligned} \log L &= n \log \theta + (\theta - 1) \log \left(\prod_{i=1}^n x_i \right) \\ &= n \log \theta + (\theta - 1) \sum_{i=1}^n \log x_i \end{aligned}$$

Taking the derivative with respect to θ ,

$$\frac{d}{d\theta} \log L = \frac{n}{\theta} + \sum_{i=1}^n \log x_i$$

2, A

Setting the derivative equal to 0,

$$\begin{aligned} \frac{n}{\theta} + \sum_{i=1}^n \log x_i &= 0 \\ \Rightarrow \frac{n}{\theta} &= - \sum_{i=1}^n \log x_i \\ \Rightarrow \frac{1}{\theta} &= \frac{- \sum_{i=1}^n \log x_i}{n} \\ \Rightarrow \theta &= \frac{-n}{\sum_{i=1}^n \log x_i} \end{aligned}$$

We need to check if this is a maximum, and by computing the second derivative,

1, B

$$\frac{d^2}{d\theta^2} \log L = -\frac{n}{\theta^2} < 0.$$

We need to still check the boundary, i.e. when $\theta \rightarrow \infty$. If $0 < x_i < 1$,

1, B

$$0 < \prod_{i=1}^n x_i < 1 \quad \Rightarrow \quad c = \left(\prod_{i=1}^n x_i \right)^{-1} > 1,$$

(if any $x_i = 0$, then the likelihood is 0, if all $x_i = 1$, then the MLE does not exist).
Writing

$$\lim_{\theta \rightarrow \infty} L(\theta|\mathbf{x}) = \lim_{\theta \rightarrow \infty} \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1} = \lim_{\theta \rightarrow \infty} \frac{\theta^n}{c^{\theta-1}},$$

and using L'Hospital's rule repeatedly,

1, C

$$\lim_{\theta \rightarrow \infty} L(\theta|\mathbf{x}) = \lim_{\theta \rightarrow \infty} \frac{\theta^n}{c^{\theta-1}} = \lim_{\theta \rightarrow \infty} \frac{n!}{(\theta-1)(\theta-2)\dots(\theta-n)c^{\theta-n-1}} = 0.$$

This shows that $\hat{\theta} = \frac{-n}{\sum_{i=1}^n \log x_i}$ is the maximum likelihood estimator of θ .

1, D

- (b) (i) An appropriate null hypothesis H_0 and an alternative hypothesis H_1 are the following: if the unknown average height of a student in the class is denoted by μ , then

meth seen ↓

$$H_0 : \mu = 175,$$

$$H_1 : \mu \neq 175.$$

1, A

- (ii) Since the variance is unknown, we can use Student's t -test. For $\alpha = 0.01$, since this is a two-sided test, and since there are $n = 18$ observations, the degrees of freedom for the t statistic is $n - 1 = 17$, we use the critical threshold $t_{17,0.995} = 2.898$.

2, B

We use the test statistic $T = \frac{175 - \bar{X}}{S^2/\sqrt{n}}$, and given the data and $\bar{x} = 176.8$ and $s^2 = 8$, our test statistic is

$$t = \frac{175 - 176.8}{\sqrt{8}/(\sqrt{18})} = \frac{-1.8}{2\sqrt{2}/(3\sqrt{2})} = \frac{-1.8}{2/3} = -2.7.$$

1, A

- (iii) Since $|t| \not> t_{17,0.995}$ ($2.7 \not> 2.898$), i.e. since the absolute value of our test statistic is not larger than the critical threshold (and so is not in the critical region), we fail to reject the null hypothesis.

1, B

Therefore, there is no evidence to suggest that the average height of the students is not 175cm, but we cannot conclude that the average height is 175cm.

1, C

- (iv) A 99% confidence interval would be

$$(176.8 - 2.898 \frac{\sqrt{8}}{\sqrt{18}}, 176.8 + 2.898 \frac{\sqrt{8}}{\sqrt{18}}) \approx (174.87, 178.73),$$

which contains 175.

1, A

- (c) One example where Student's two-sample t -test can be used is to show that the average heights of two groups of people are **not** equal.

seen ↓

1, C

We need to assume that the heights of each group (1) follow a normal distribution, and (2) are independent.

1, B

We also need to assume (3) that the variances of the heights of the two groups are equal, even though this variance is unknown.

1, C

- (d) Under the null hypothesis a p -value P_i (when considered as a random variable) follows a $U(0, 1)$ distribution.

unseen ↓

1, B

Therefore, for each p -value P_i ,

$$P(P_i \leq 0.05) = 0.05.$$

1, C

Let X be the random variable that is the number of p -values less than 0.05. Then X follows a binomial distribution with parameters $n = 200$ and $q = 0.05$ (using q to avoid confusion with the p -values).

1, D

Since $E[X] = nq$, the expected value of X is $200 \cdot 0.05 = 10$. Therefore, we expect 10 p -values to be less than $\alpha = 0.05$.

1, D

6. (a) (i) Since x_i , β_0 and β_1 are all constants, using the property of the variance,

seen ↓

$$\text{Var}(Y_i) = \text{Var}(\beta_0 + \beta_1 x_i + \epsilon_i) = \text{Var}(\epsilon_i) = \sigma^2,$$

as required.

1, A

- (ii) Since

meth seen ↓

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

one can show

1, A

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

as required.

1, A

- (iii) Using the results in Part (i) and Part (ii),

unseen ↓

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

3, B

where we have used the independence of the Y_i , so $\text{Cov}(Y_i, Y_j) = 0$ if $i \neq j$, which reduces the sum to only be the i th term.

1, C

- (iv) One example would be to find the linear relationship between two physical quantities, such as altitude and the boiling point of water (perhaps after a transformation, such as taking the logarithm).

sim. seen ↓

1, B

Finding the estimates for the values of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ would allow us to predict a value of y given a new observation x . In this example, that would mean being able to predict the boiling point of water for an altitude not previously measured.

2, D

(b) The probability mass function for X_i is

sim. seen ↓

$$f(x_i|\theta) = \theta(1 - \theta)^{x_i-1},$$

and if we assume that the random variables $X_1, X_2, \dots, X_n$ are independent, then

1, A

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f(x_i|\theta) \\ &= \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} \\ &= \theta^n (1 - \theta)^{\sum_{i=1}^n (x_i-1)} \\ &= \theta^n (1 - \theta)^{(n\bar{x}-n)}. \end{aligned}$$

1, A

The prior distribution for θ is

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \quad \text{for } \theta \in (0, 1).$$

The posterior p.d.f. can be computed as

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\ &= \theta^n (1 - \theta)^{n\bar{x}-n} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha+n-1} (1 - \theta)^{\beta+n\bar{x}-n-1} \\ &\propto \theta^{\alpha+n-1} (1 - \theta)^{\beta+n\bar{x}-n-1}, \end{aligned}$$

for $\theta \in (0, 1)$.

2, B

Therefore, we see that the posterior distribution (once given the appropriate scaling constant) is a $\text{Beta}(\alpha + n, \beta + n\bar{x} - n)$ distribution, i.e.

$$\pi(\theta|\mathbf{x}) = \frac{\Gamma(\alpha + \beta + n\bar{x})}{\Gamma(\alpha + n)\Gamma(\beta + n\bar{x} - n)} \theta^{\alpha+n-1} (1 - \theta)^{\beta+n\bar{x}-n-1},$$

for $\theta \in (0, 1)$.

1, C

(c) The model does not fit the data well

sim. seen ↓

because the residuals are in a U-shape.

1, B

1, C

(d) (i) Simpson's paradox.

sim. seen ↓

1, A

(ii) If the subgroups have been split unevenly between the treatments, or if the subgroups are of very different sizes (although that is not the case in this example), then this can provide conditions that allow Simpson's paradox to occur.

sim. seen ↓

1, D

(iii) In this example, it is possible that Group A consists of patients who have a very high chance of developing the disease, and so a larger proportion of patients in Group A were given the drug, while Group B consists of patients who have a lower chance of developing the disease, and so more were given the placebo (the difference between the groups could be due to age or gender, for example). (Other reasonable explanations are also acceptable.)

sim. seen ↓

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

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

Please record below, some brief but non-trivial comments for students about how well (or otherwise) the questions were answered. For example, you may wish to comment on common errors and misconceptions, or areas where students have done well. These comments should note any errors in and corrections to the paper. These comments will be made available to students via the MathsCentral Blackboard site and should not contain any information which identifies individual candidates. Any comments which should be kept confidential should be included as confidential comments for the Exam Board and Externals. If you would like to add formulas, please include a sperate pdf file with your email.

ExamModuleCode	QuestionNumber	Comments for Students
MATH40005 Probability & Statistics	1 a)	Generally well done, though often not followed through with an actual calculation to corroborate the probability estimate assertion.
MATH40005 Probability & Statistics	1 b) i)	Generally well done - but often the fact that there were 7 rolls of the die was missed
MATH40005 Probability & Statistics	1 b) ii)	Generally well done - but often the fact that the question asked for the cardinality of the powerset was missed.
MATH40005 Probability & Statistics	1 b) iii)	Often well done - but the probability space elements, particularly, the proabability measure, were not always clarified sufficiently
MATH40005 Probability & Statistics	1 c)	Generally well done - but in a few cases the definitions of independence and mutually exclusive were confused
MATH40005 Probability & Statistics	1 d)	Sometimes well done - but often only the answer was provided without justification
MATH40005 Probability & Statistics	1 e)	Sometimes well done - but often only the answer was provided without justification and in a few cases 'outcome' was misinterpreted

MATH40005
Probability & Statistics 1 f)

Sometimes well done - but very often without justification (which was easy to provide)

In part (a) many students tried to answer a totally different question from the one being asked. It was a good test of their understanding of conditional probability.

MATH40005
Probability & Statistics Q2

In part (b) there were some good answers but also some misconceptions: note that $E(X/Y)$ is not in general equal to $E(X)/E(Y)$. This was a good question as it tested a number of different aspects of their understanding.

Part (c) was done reasonably well by those who were familiar with the definitions

Part (d) was done poorly on the whole. Many students mistakenly thought that the independence of X and Y implied that the mgf of Z is the product of the separate mgfs of X and $\sqrt{Y}$.

MATH40005
Probability & Statistics Q3

Question 3(a). i was well answered, only a few students did not specify the limits. In question 3(a). ii some struggled to provide a whole justification on why the new variable is continuous, not showing that the cumulative can be differentiated. Some students only provided the value of median and mean on question 3(a). iii without any mathematical proof.

Many students managed to answer Question 3(b). Some forgot about the justification on why this result is useful.

Questions 3(c). i and ii were generally answered well with no problems in general, just some minor calculation errors. On the other hand, question 3(c) iii was the one that students struggled the most and a substantial number of students left it blank.

MATH40005
Probability & Statistics Q4

Part (a): Some students found this question challenging, especially the last two parts as they did not correctly identify an unbiased estimator. Part (b): Almost half of the students answered this question incorrectly by assuming a Normal distribution for the data. Part (c): This part was successfully answered by the majority of the students.

MATH40005
Probability & Statistics

Q5

The derivation of the MLE in part (a) was generally correctly done but many students forgot to check the boundaries. Most of the students correctly answered part (b); some students did not explicitly compute the test statistic. In part (c) most students provided an example of the use of the two sample test and stated the correct assumptions; a common mistake was to state that the test could be used to show that the average of two samples are equal (instead of different). Question (d) was the most challenging one; only a few students provided the correct answer with full justification.

MATH40005
Probability & Statistics

Q6

6(a)(i), 6(a)(ii), 6(c) and (d) were generally answered well, although a fair number of students did not use the property of the variance operator in 6(a)(i) or failed to provide justification. Many students struggled with 6(a)(iii), but this could be easily solved using the covariance identities introduced in the problem sheets/problem-based learning sheets. Many students also chose to omit 6(a)(iv); there were two real-world examples of linear regression in the notes, e.g. boiling point of water vs air pressure, and the key use of linear regression is that it allows one to use data to predict the value of a y-value (air pressure) given another x-value (boiling point of water). The Bayesian question 6(b) was generally answered well.