

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2022

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

Probability and Statistics

Date: 12 May 2022

Time: 09:00 – 12:00 (BST)

Time Allowed: 3:00 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 AS 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 the sample space $\Omega = \{A, B, C, D\}$.

- (i) Provide an example, with justification, of a sigma-algebra on Ω which is neither the trivial nor the power-sigma algebra. (2 marks)
- (ii) Now suppose that $\mathcal{F} = \mathcal{P}(\Omega)$, i.e. $\mathcal{F}$ is the power-sigma algebra on Ω . Give an example, with justification, of a discrete random variable X on $(\Omega, \mathcal{F})$. (3 marks)
- (iii) Again suppose that $\mathcal{F} = \mathcal{P}(\Omega)$, i.e. $\mathcal{F}$ is the power-sigma algebra on Ω . Give an example, with justification, of a discrete random variable Y which has a discrete uniform distribution. Remember that you need to define a suitable probability measure on $(\Omega, \mathcal{F})$. (5 marks)

(b) Consider the following identity: For $n \in \mathbb{N}$, we have

$$n \cdot 4^{n-1} = \sum_{k=0}^n \binom{n}{k} 3^k (n-k). \quad (1)$$

- (i) Prove the identity (1) using direct calculations or a proof by induction.
Hint: You might find the binomial theorem helpful. (3 marks)
- (ii) Prove the identity (1) using a "story-proof", where you provide a suitable interpretation of the terms on the left-hand-side and the right-hand-side of the equation. (7 marks)

(Total: 20 marks)

2. (a) For $m, n \in \mathbb{N}$, consider random variables $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ and constants $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{R}$. Show that

$$\text{Cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j),$$

assuming the covariances exist. (3 marks)

- (b) Consider two random variables X and Y with $E(X) = 2$, $\text{Var}(X) = 5$, $E(Y) = -4$, $\text{Var}(Y) = 7$, $\text{Cov}(X, Y) = 3$.

- (i) Define the random variable $Z = 3X - 2Y$. Find $E(Z)$ and $\text{Var}(Z)$. (5 marks)
(ii) Find $\text{Cov}(2(X + Y), 3(X - Y))$. (4 marks)

- (c) Let X denote a real-valued, continuous random variable with continuous probability density function f_X . Let Y denote a discrete random variable with $\text{Im}Y = \{1, 2, \dots, 10\}$. Suppose that X and Y are independent. Define a new random variable by $Z = XY$.

- (i) Show that the random variable Z is continuous and find its probability density function f_Z . Remember to show that your f_Z is a probability density function. (5 marks)
(ii) Consider the same setting as above except that Y may take the value 0, with positive probability, i.e. $\text{Im}Y = \{0, 1, 2, \dots, 10\}$. Find the cumulative distribution function of Z . (3 marks)

(Total: 20 marks)

3. (a) The newspaper Guardian reported in an article entitled "The rise of lateral flow tests: are these 'heroes' of the pandemic here to stay?" on 7th January 2022 the following:

"As case numbers fell in 2021, lateral flow tests also came under criticism for delivering false positives. This reflects a classic statistical conundrum: if you test positive, the chances that you really have Covid depends on not only the quality of the test but also the current prevalence of the virus. For instance, if an LFT delivers a false positive result for one in 1,000 tests at a time when just one in 1,000 people have Covid, the chances that someone with a positive [result] is truly infected are only 50%. But if the prevalence is one in 10 people, as it was in London last week, that person can be 99% sure they are infected."

Note that LFT stands for lateral flow test.

Can you verify the statement above using results from the lectures? Did the journalist make any implicit assumptions? (8 marks)

- (b) Consider two independent continuous random variables X, Y which are uniformly distributed on $[0, 1]$, i.e. their probability density functions are given by $f(x) = 1$ for $x \in [0, 1]$ and 0 otherwise.

Find

$$P\left(\max\{X, Y\} > \frac{3}{4} \mid \min\{X, Y\} > \frac{1}{3}\right).$$

(6 marks)

- (c) Consider two jointly continuous random variables X and Y with joint probability density function given by

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

for a constant $c \in \mathbb{R}$.

- (i) Show that $c = \frac{24}{7}$ (1 mark)
- (ii) Find the marginal density function of X . (2 marks)
- (iii) Find the conditional expectation $E(Y|X = x)$ and specify for which values of x it is well-defined. (3 marks)

(Total: 20 marks)

4. (a) Suppose that $X_1, X_2, \dots, X_n$ are independent and identically distributed random variables, all following the same distribution with known variance σ^2 . The mean of this distribution is unknown and is denoted by θ . We decide to define an estimator of θ as

$$\hat{\Theta} = \frac{1}{n+1} \sum_{i=1}^n X_i.$$

Showing all working and clearly stating any results used:

- (i) Compute $b_{\theta}(\hat{\Theta})$, the bias of $\hat{\Theta}$. (2 marks)
- (ii) Compute the mean squared error of $\hat{\Theta}$. (3 marks)
- (b) Suppose a pharmaceutical company plans to conduct a clinical trial for a candidate vaccine that prevents severe symptoms for a particular disease. For the trial, the company will recruit n volunteers. Each volunteer will be given the vaccine and then monitored over the next three months to check whether or not they have developed severe symptoms for the disease. The company would like to use this trial to estimate the true proportion p of vaccinated individuals who would develop severe symptoms (supposing the vaccine were made available to everyone, and that most people would be exposed to the disease within three months). What value of n should be chosen so that p could be estimated to within $\delta = 0.02$ with confidence 0.95? Provide a careful argument to justify your answer. (6 marks)
- (c) Suppose the continuous random variable Y follows a $\text{Beta}(\alpha, \beta)$ distribution with parameters $\alpha > 0$ and $\beta > 0$. Recall that the probability density function of Y is therefore

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}, & \text{if } y \in (0, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Suppose we can further assume that $\beta = 1$ and $\alpha > 1$, although the precise value of α is unknown. Show that the median of Y is $m = 2^{-1/\alpha}$. Hint: you may use the identity $\Gamma(x+1) = x\Gamma(x)$, for all $x > 0$. (3 marks)

- (d) Suppose that the random variables $Z_1, Z_2, \dots, Z_n$ represent the daily changes (increases or decreases) in the share price of a company for the past n days, and these random variables are assumed to be independent and identically distributed according to a normal distribution with unknown mean μ and unknown variance σ^2 . For $n = 20$ days, the random variables are observed as $z_1, z_2, \dots, z_n$, and their sample mean is computed as $\bar{z} = 10.5$, where the units are British pounds.
 - (i) Suppose that the sample variance of $z_1, z_2, \dots, z_n$ is computed as $s^2 = 580$. Construct a 99% confidence interval for the unknown mean μ , and calculate the realised confidence interval using the given values of $\bar{z}$ and s^2 . Note that Tables 1 and 2 on pages 9 and 10 may be helpful. (4 marks)
 - (ii) Your colleague claims that this data shows the share price of the company has generally been increasing for the past n days. Given the confidence interval in Part (i), comment on this claim. (2 marks)

(Total: 20 marks)

5. (a) Suppose one defines the null hypothesis of an experiment to be H_0 .
- (i) Define α and β , the Type I and Type II errors, respectively, in terms of probabilities relating to H_0 . (2 marks)
 - (ii) Is it always possible to maximise statistical power without affecting the Type I and/or Type II errors? Briefly justify your answer. (2 marks)
 - (iii) In the real-world example of a court case where the null hypothesis is that the defendant is innocent, define the Type I and Type II errors in this context. (2 marks)
- (b) The coach of a basketball team would like to see if the current players are taller than in previous years. Suppose the heights of the players are the independent and identically distributed random variables $X_1, X_2, \dots, X_n$, where $n = 15$, and these random variables are assumed to follow a normal distribution with unknown mean θ and known variance σ^2 . The coach would like to see if $\theta > 185$, where the units are cm, and plans to measure the heights of the players to test this hypothesis using a significance threshold of $\alpha = 0.05$.
- (i) Write down the null and alternative hypotheses for this experiment. (2 marks)
 - (ii) Suppose the heights $x_1, x_2, \dots, x_{15}$ are measured and the sample mean is computed as $\bar{x} = 187$ (in cm). Assuming that $\sigma = 3$, what can be concluded about θ , given the hypotheses in (i) and the significance threshold α above? Note that Tables 1 and 2 on pages 9 and 10 may be helpful. (4 marks)
- (c) Suppose a research lab is trying to identify genetic mutations that are significantly associated with a particular disease that is assumed to be hereditary. In order to do this, the genomes of a healthy cohort of volunteers is sequenced along with the genomes of a cohort of volunteers who have been diagnosed with the disease, and statistical tests of association are carried out on 1000 different genetic mutations, resulting in 1000 separate p -values, with the two smallest being 5×10^{-6} and 2×10^{-5} (and all other p -values are larger). The director of the lab is pleased, because the significance threshold had been pre-specified as $\alpha = 0.001$.
- (i) Would the director be correct in concluding that both of the two smallest p -values are significant? Briefly justify your answer. (2 marks)
 - (ii) Regardless of the answer in (i), what could be changed so that exactly one of these two p -values can be declared significant? (2 marks)
- (d) Suppose the heights of n pairs of siblings are measured as $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where $n = 50$, and the following quantities are provided:

$$\bar{x} = 5.34, \quad \bar{y} = 4.72, \quad \frac{1}{n} \sum_{i=1}^n x_i^2 = 38.86, \quad \frac{1}{n} \sum_{i=1}^n y_i^2 = 50.92, \quad \frac{1}{n} \sum_{i=1}^n x_i y_i = 37.62.$$

Compute the sample correlation r_{XY} . (4 marks)

(Total: 20 marks)

6. (a) Suppose that the length of n posts on an internet discussion forum are the independent random variables $X_1, X_2, \dots, X_n$ which are each assumed to follow the distribution with probability density function

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log x - \theta)^2\right), & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where σ^2 is a known constant, $\log x$ is the natural logarithm (with base e), and the parameter $\theta \in (-\infty, \infty)$ is unknown. Given that the random variables are observed as $x_1, x_2, \dots, x_n$ (all values greater than 0), find $\hat{\theta}$, the maximum likelihood estimate of θ , and $\hat{\Theta}$, the maximum likelihood estimator of θ .

(6 marks)

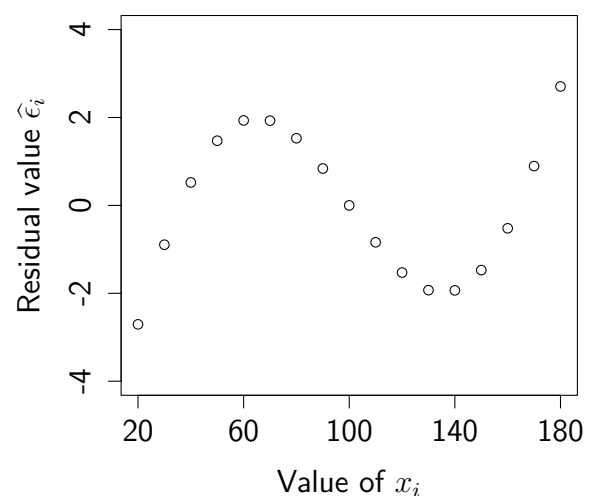
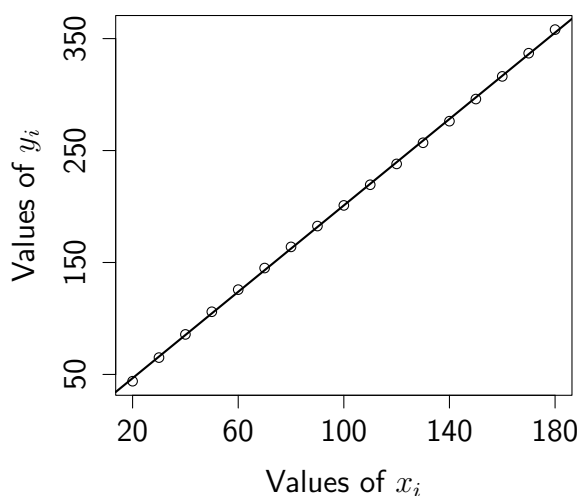
- (b) Suppose that the following data are recorded from a small observational study:

$$\{21, 54, 32, 98, 10\}.$$

For each of the following samples, state whether or not they are bootstrap samples of this data, and provide justification.

- (i) $\{54, 21, 21, 10, 10\}$ (1 mark)
(ii) $\{21, 54, 32, 98, 76\}$ (1 mark)

- (c) Suppose a scientist records the following data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ for two quantities X and Y which are believed to be linearly related. The scientist fits a simple linear regression model to the data and obtains the fit and residual plot shown in the figures below, along with an R^2 value of $R^2 = 0.9997$.

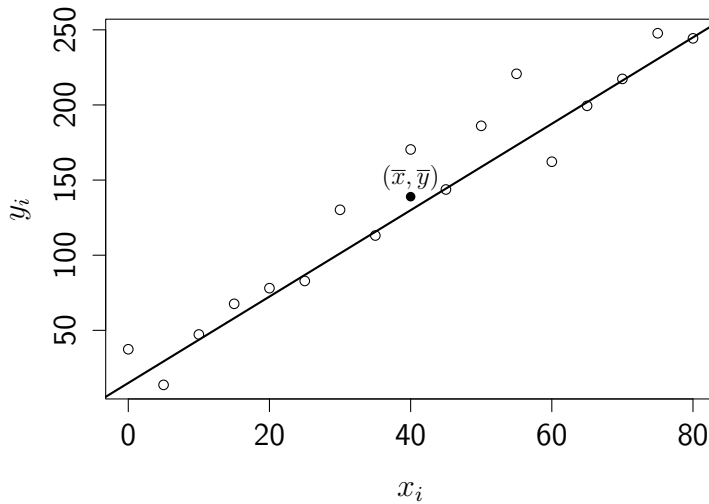


Given this information, briefly assert and justify whether or not this model fits the data well. (2 marks)

- (d) One of your friends collects data $(x_1, z_1), (x_2, z_2), \dots, (x_n, z_n)$, and after using the transformation $y_i = \log z_i$, believes that the data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are linearly related, so fits the simple linear regression model

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

to the data. After calculating the maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 , respectively, your friend plots the data and the regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ as shown below. A second friend looks at the plot and, after filling in and labelling the point $(\bar{x}, \bar{y})$ as a solid black point on the plot, declares a mistake has been made in the calculations of $\hat{\beta}_0$ and $\hat{\beta}_1$. Note that, as usual, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.



Given the plot above, your two friends ask you to decide if the fitted linear regression line is correct or not. Provide your answer, with full justification. (4 marks)

- (e) Suppose the daily rainfall for n days in a particular city is modelled by the random variables $X_1, X_2, \dots, X_n$ which are assumed to be independent and to follow the distribution with probability density function

$$g(x) = \begin{cases} \mu\theta \exp(\theta + \mu x - \theta e^{\mu x}), & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu > 0$ is a known constant and $\theta > 0$ is an unknown parameter. Following a Bayesian approach, assume that the prior distribution of θ has probability density function

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

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 posterior distribution by stating the family of distributions to which it belongs as well as its parameter values. (6 marks)

(Total: 20 marks)

Table 1: $P(Z < z)$ where $Z \sim N(0, 1)$ for values of z between 1.00 and 2.99. Note that this is a reduced version of the table in the notes. One reads the table as $P(Z < z_r + z_c) = p_{r,c}$, where 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, $P(Z < 1.02) = 0.8461$.

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

Table 2: Values of t for $P(T < t)$, where T has Student's t -distribution with ν degrees of freedom, for $\nu \in \{1, 2, \dots, 20\}$. 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, $t_{3,0.60} = 0.277$.

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

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2022

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

MATH40005

Probability and Statistics (Solutions)

Setter's signature

.....

Checker's signature

.....

Editor's signature

.....

1. (a) (i) Various examples are possible. E.g.:

meth seen ↓

- $\mathcal{F} = \{\emptyset, \Omega, \{A\}, \{A\}^c\} = \{\emptyset, \Omega, \{A\}, \{B, C, D\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{B\}, \{B\}^c\} = \{\emptyset, \Omega, \{B\}, \{A, C, D\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{C\}, \{C\}^c\} = \{\emptyset, \Omega, \{C\}, \{A, B, D\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{D\}, \{D\}^c\} = \{\emptyset, \Omega, \{D\}, \{A, B, C\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{A, B\}, \{A, B\}^c\} = \{\emptyset, \Omega, \{A, B\}, \{C, D\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{A, C\}, \{A, C\}^c\} = \{\emptyset, \Omega, \{A, C\}, \{B, D\}\}.$
- $\mathcal{F} = \{\emptyset, \Omega, \{A, D\}, \{A, D\}^c\} = \{\emptyset, \Omega, \{A, D\}, \{B, C\}\}.$

For each case, we note that $\emptyset \in \mathcal{F}$, $\mathcal{F}$ is closed under complements, i.e. if $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$, and $\mathcal{F}$ is closed under countable union, i.e. for any $E_1, E_2, \dots \in \mathcal{F}$, we have that $\cup_i E_i \in \mathcal{F}$.

2, B

[Only one example is needed for full marks. 2 marks for a correct $\mathcal{F}$ and 1 mark for the justification.]

meth seen ↓

- (ii) Various examples exist. We give the details of one example here: We define the random variable $X : \Omega \rightarrow \mathbb{R}$ as $X(\omega) = 1$ for all $\omega \in \Omega$. We note that $\text{Im}X = \{1\}$ is finite, hence countable. Also, $X^{-1}(1) = \Omega \in \mathcal{F}$ and $X^{-1}(x) = \emptyset \in \mathcal{F}$ for all $x \in \mathbb{R} \setminus \{1\}$. Hence X is a discrete random variable on $(\Omega, \mathcal{F})$.

3, B

- (iii) Various examples/notations exist. We give the details of one example here: We define the function $Y : \Omega \rightarrow \mathbb{R}$ such that

$$Y(A) = 1, \quad Y(B) = 2, \quad Y(C) = 3, \quad Y(D) = 4.$$

We note that $\text{Im}Y = \{1, 2, 3, 4\}$ is finite, hence countable. Also, $Y^{-1}(1) = \{A\} \in \mathcal{F}$, $Y^{-1}(2) = \{B\} \in \mathcal{F}$, $Y^{-1}(3) = \{C\} \in \mathcal{F}$, $Y^{-1}(4) = \{D\} \in \mathcal{F}$, and $Y^{-1}(x) = \emptyset \in \mathcal{F}$ for all $x \in \mathbb{R} \setminus \{1, 2, 3, 4\}$. Hence Y is a discrete random variable on $(\Omega, \mathcal{F})$.

2, B

We want Y to have a discrete uniform distribution. We consider a discrete uniform distribution on $S := \text{Im}Y = \{1, 2, 3, 4\}$. We assign the probabilities $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 1/4$ to each of the four outcomes.

As in the lecture, we define a probability measure $P : \mathcal{F} \rightarrow [0, 1]$ as follows. $P(E) = \sum_{i: s_i \in E} \pi_i = \text{card}(E)/4$ for all $E \in \mathcal{F}$.

1, C

According to lectures, this is indeed a valid probability measure, and we get that

$$P(Y = 1) = P(Y = 2) = P(Y = 3) = P(Y = 4) = \frac{1}{4}.$$

Hence Y follows a discrete uniform distribution on $\{1, 2, 3, 4\}$.

2, C

- (b) (i) We recall the binomial theorem: For $n \in \mathbb{N}$, $x, y \in \mathbb{R}$, we have

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

We give a proof by direct calculation (an alternative proof by induction can use the same ideas): Starting from the left-hand-side, and applying the binomial theorem, we have, for $n \in \mathbb{N}$,

$$n \cdot 4^{n-1} = n \cdot (3 + 1)^{n-1} = n \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} 3^k 1^{n-1-k} = n \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} 3^k.$$

We note that

$$n \cdot \binom{n-1}{k} = n \cdot \frac{(n-1)!}{(n-1-k)!k!} = \frac{n!}{(n-k)!k!} \cdot (n-k) = \binom{n}{k} \cdot (n-k).$$

Hence, we have that

$$n \cdot 4^{n-1} = \sum_{k=0}^{n-1} \binom{n}{k} 3^k (n-k) = \sum_{k=0}^n \binom{n}{k} \cdot 3^k (n-k),$$

where we used the fact that the (last) term in the sum, when $k = n$ is equal to 0.

3, B

- (ii) [Various different stories/interpretations are possible, we provide one example of a correct answer] Suppose n friends go for a night out. They decide that one of them should stay sober to drive the car/van. Each person can select one of four drinks: red wine, white wine, beer or water. The driver can only drink water. We now count the number of possible drinks orders, where at least one is water.

Let us consider the left hand side first: We observe that the number of possibilities to choose one driver from the n people, which is given by $\binom{n}{1} = n$, is multiplied by the number of possibilities that the remaining $n - 1$ people, independently, pick one out of 4 possible drinks, which is given by $\binom{4}{1}^{n-1} = 4^{n-1}$.

3, D

On the right hand side, we can first select k out of n people to choose one out of the three alcoholic drink options, the number of possibilities is given by $\binom{n}{k} \binom{3}{1}^k = \binom{n}{k} 3^k$. Next, we choose the driver, i.e. one of the remaining $n - k$ people, who all also pick the water option as their drink. Here we have $\binom{n-k}{1} \binom{1}{1} = n - k$ possibilities. Summing the product of these two terms over all possible k s, we observe that the situation is identical to the one described on the left hand side.

4, D

2. (a) For $m, n \in \mathbb{N}$, consider random variables $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ and constants $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{R}$. We have

meth seen ↓

$$\begin{aligned}
 & \text{Cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right) \\
 & \stackrel{\text{definition of covariance}}{=} \mathbb{E} \left(\sum_{i=1}^m a_i X_i \sum_{j=1}^n b_j Y_j \right) - \mathbb{E} \left(\sum_{i=1}^m a_i X_i \right) \mathbb{E} \left(\sum_{j=1}^n b_j Y_j \right) \\
 & = \mathbb{E} \left(\sum_{i=1}^m \sum_{j=1}^n a_i b_j X_i Y_j \right) - \mathbb{E} \left(\sum_{i=1}^m a_i X_i \right) \mathbb{E} \left(\sum_{j=1}^n b_j Y_j \right) \\
 & \stackrel{\text{linearity of expectation}}{=} \sum_{i=1}^m \sum_{j=1}^n a_i b_j \mathbb{E}(X_i Y_j) - \sum_{i=1}^m a_i \mathbb{E}(X_i) \sum_{j=1}^n b_j \mathbb{E}(Y_j) \\
 & = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \mathbb{E}(X_i Y_j) - \sum_{i=1}^m \sum_{j=1}^n a_i b_j \mathbb{E}(X_i) \mathbb{E}(Y_j) \\
 & = \sum_{i=1}^m \sum_{j=1}^n a_i b_j [\mathbb{E}(X_i Y_j) - \mathbb{E}(X_i) \mathbb{E}(Y_j)] \\
 & \stackrel{\text{definition of covariance}}{=} \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j),
 \end{aligned}$$

assuming the covariances exist.

[1 mark for justifications, 2 marks for calculation]

3, A

- (b) (i) Using the linearity of the expectation, we get

meth seen ↓

$$\mathbb{E}(Z) = 3\mathbb{E}(X) - 2\mathbb{E}(Y) = 3 \cdot 2 - 2 \cdot (-4) = 14.$$

Adapting the result in 2(a), we get that

2, A

$$\begin{aligned}
 \text{Cov}(a_1 X + a_2 Y, a_1 X + a_2 Y) &= \text{Var}(a_1 X + a_2 Y) \\
 &= a_1^2 \text{Var}(X) + a_2^2 \text{Var}(Y) + 2a_1 a_2 \text{Cov}(X, Y).
 \end{aligned}$$

Hence

$$\begin{aligned}
 \text{Var}(Z) &= \text{Var}(3X - 2Y) = 3^2 \text{Var}(X) + (-2)^2 \text{Var}(Y) + 2 \cdot 3 \cdot (-2) \text{Cov}(X, Y) \\
 &= 9 \cdot 5 + 4 \cdot 7 - 12 \cdot 3 = 37.
 \end{aligned}$$

3, A

- (ii) Using the formula provided in 2(a), we deduce that

$$\begin{aligned}
 \text{Cov}(2(X + Y), 3(X - Y)) &= 2 \cdot 3 \cdot \text{Cov}((X + Y), (X - Y)) \\
 &= 6 \cdot \text{Cov}(X + Y, X - Y),
 \end{aligned}$$

where

$$\begin{aligned}
 \text{Cov}(X + Y, X - Y) &= \text{Cov}(X, X) - \text{Cov}(X, Y) + \text{Cov}(Y, X) - \text{Cov}(Y, Y) \\
 &= \text{Var}(X) - \text{Var}(Y) = 5 - 7 = -2,
 \end{aligned}$$

where we used the symmetry of the covariance, i.e. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

So, altogether, we get that $\text{Cov}(2(X + Y), 3(X - Y)) = -12$.

4, A

- (c) (i) We first derive the cumulative distribution function (cdf) of Z which is given as follows. Let $z \in \mathbb{R}$, then

unseen ↓

$$F_Z(z) = P(Z \leq z) = P(XY \leq z) \stackrel{\text{Law of total probability}}{=} \sum_{y=1}^{10} P(XY \leq z | Y = y) P(Y = y)$$

1, A

$$= \sum_{y=1}^{10} P(Xy \leq z | Y = y) P(Y = y) \stackrel{y \geq 0}{=} \sum_{y=1}^{10} P\left(X \leq \frac{z}{y} \middle| Y = y\right) P(Y = y)$$

$$\stackrel{\text{Independence of } X \text{ and } Y}{=} \sum_{y=1}^{10} P\left(X \leq \frac{z}{y}\right) P(Y = y) = \sum_{y=1}^{10} F_X(z/y) P(Y = y),$$

where we denote by F_X the cdf of X . Since X is a continuous random variable with continuous density f_X , we know that F_X and hence also F_Z is differentiable. Differentiating F_Z leads to

1, C

$$\frac{d}{dz} F_Z(z) = \frac{d}{dz} \sum_{y=1}^{10} F_X(z/y) P(Y = y) = \sum_{y=1}^{10} f_X(z/y) \frac{1}{y} P(Y = y) =: f_Z(z),$$

for $z \in \mathbb{R}$, where we applied the fundamental theorem of calculus and the chain rule for differentiation.

1, C

We summarise that Z is a random variable whose cdf is given by $F_Z(z) = \int_{-\infty}^z f_Z(u) du$, for all $z \in \mathbb{R}$, where $f_Z : \mathbb{R} \rightarrow \mathbb{R}$ such that f_Z is a probability density function, i.e. $f_Z(u) = \sum_{y=1}^{10} f_X(u/y) \frac{1}{y} \geq 0$ for all $u \in \mathbb{R}$ [since $f_X(x) \geq 0$ for all $x \in \mathbb{R}$ and $y \in \{1, \dots, 10\}$] and

$$\int_{-\infty}^{\infty} f_Z(u) du = \int_{-\infty}^{\infty} \sum_{y=1}^{10} f_X(u/y) \frac{1}{y} P(Y = y) du = \sum_{y=1}^{10} P(Y = y) \int_{-\infty}^{\infty} f_X(u/y) \frac{1}{y} du$$

$$\stackrel{\text{change of variable } u/y=v}{=} \underbrace{\sum_{y=1}^{10} P(Y = y)}_{=1} \underbrace{\int_{-\infty}^{\infty} f_X(v) dv}_{=1} = 1,$$

where we used that the density f_X integrates to 1 and that the probability mass function of Y sums up to 1.

2, C

- (ii) We adapt the above computations to include the case when $Y = 0$. Let $z \in \mathbb{R}$, then

unseen ↓

$$F_Z(z) = P(Z \leq z) = P(XY \leq z)$$

$$\stackrel{\text{Law of total probability}}{=} P(XY \leq z | Y = 0) P(Y = 0) + \sum_{y=1}^{10} P(XY \leq z | Y = y) P(Y = y)$$

1, A

$$\stackrel{(i)}{=} P(0 \leq z) P(Y = 0) + \sum_{y=1}^{10} F_X(z/y) P(Y = y).$$

Hence, for $z \geq 0$, we have $F_Z(z) = P(Y = 0) + \sum_{y=1}^{10} F_X(z/y) P(Y = y)$, and, for $z < 0$, $F_Z(z) = \sum_{y=1}^{10} F_X(z/y) P(Y = y)$.

2, C

3. (a) We use the law of total probability and Bayes' theorem to verify this result. We denote by C the event that a person has Covid and by TP the event that a person tests positive.

meth seen ↓

We are given the information that

1. Case 1: $P(C) = 1/1000 = 0.001$, which implies that $P(C^c) = 0.999$,

2. Case 2: $P(C) = 1/10 = 0.1$, which implies that $P(C^c) = 0.9$,

and that $P(TP|C^c) = 1/1000 = 0.001$.

Using Bayes' theorem, we find that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)}.$$

The term $P(TP)$ can be derived using the law of total probability:

$$P(TP) = P(TP|C)P(C) + P(TP|C^c)P(C^c).$$

2, A

The conditional probability $P(TP|C)$ is not provided, but is needed for the computations. Possibly, the journalist made the implicit and idealised assumption that $P(TP|C) = 1$, i.e. that the conditional probability that the LFT is positive given the person has Covid is 1.

2, B

Suppose now that $P(TP|C) = 1$.

Case 1:

By the law of total probability, we get that

$$\begin{aligned} P(TP) &= P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.001 + 0.001 \cdot 0.999 \\ &= 0.001 + \frac{999}{1000000} = \frac{1999}{1000000} = 0.001999. \end{aligned}$$

Using Bayes' theorem, we get that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)} = \frac{0.001}{0.001999} = \frac{0.001}{0.001999} \approx 0.5.$$

Hence, the probability that somebody has Covid given they have a positive test is approximately 50% when 1 in 1000 in the population is infected.

2, A

Case 2:

By the law of total probability, we get that

$$\begin{aligned} P(TP) &= P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.1 + 0.001 \cdot 0.9 \\ &= \frac{1009}{10000} = 0.1009. \end{aligned}$$

Using Bayes' theorem, we get that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)} = \frac{0.1}{0.1009} \approx 0.99.$$

Hence, the probability that somebody has Covid given they have a positive test is approximately 99% when 1 in 10 in the population is infected.

2, A

[Alternative approaches: Students might consider the approach that the unknown probability is given as $P(TP|C) = x$ for an x such that $P(C|TP) = 0.5$ in Case 1 or $P(C|TP) = 0.99$ in Case 2. We remark that there is no x which satisfies both $P(C|TP) = 0.5$ in Case 1 and $P(C|TP) = 0.99$ in Case 2 simultaneously. One of the two equations holds only approximately as the following computations show. Full marks should be awarded for correct alternative solutions or if a student shows in full detail that the two equations cannot hold exactly simultaneously. These alternative approaches are computationally heavier.

Alternative approach 1:

Case 1:

By the law of total probability, we get that

$$\begin{aligned} P(TP) &= P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.001x + 0.001 \cdot 0.999 \\ &= \frac{1000x + 999}{1000000} = 0.001x + 0.000999. \end{aligned}$$

Using Bayes' theorem, we get that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)} = \frac{0.001x}{0.001x + 0.000999} \stackrel{!}{=} 0.5,$$

which is the case if and only if

$$0.001x = 0.5(0.001x + 0.000999) \iff x = \frac{0.5 \cdot 0.000999}{0.001 - 0.5 \cdot 0.001} = \frac{0.0004995}{0.0005} = 0.999.$$

Case 2: Now assume that $P(TP|C) = x = 0.999$. By the law of total probability, we get that

$$\begin{aligned} P(TP) &= P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.1x + 0.001 \cdot 0.9 \\ &= 0.1x + 0.0009 = 0.1 \cdot 0.999 + 0.0009 = 0.1008. \end{aligned}$$

Using Bayes' theorem, we get that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)} = \frac{x \cdot 0.1}{0.1x + 0.0009} = \frac{0.0999}{0.1008} \approx 0.99.$$

Alternative approach 2:

Case 2: By the law of total probability, we get that

$$P(TP) = P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.1x + 0.001 \cdot 0.9 = 0.1x + 0.0009.$$

Using Bayes' theorem, we get that

$$P(C|TP) = \frac{P(TP|C)P(C)}{P(TP)} = \frac{x \cdot 0.1}{0.1x + 0.0009} \stackrel{!}{=} 0.99,$$

which is the case if and only if

$$0.1x = 0.99(0.1x + 0.0009) \iff x = \frac{0.99 \cdot 0.0009}{0.1 - 0.99 \cdot 0.1} = 0.891.$$

Case 1: Now assume that $P(TP|C) = x = 0.891$. By the law of total probability, we get that

$$\begin{aligned} P(TP) &= P(TP|C)P(C) + P(TP|C^c)P(C^c) = 0.001x + 0.001 \cdot 0.999 \\ &= 0.001 \cdot 0.891 + 0.001 \cdot 0.999 = 0.00189. \end{aligned}$$

Using Bayes' theorem, we get that

$$\begin{aligned} P(C|TP) &= \frac{P(TP|C)P(C)}{P(TP)} = \frac{0.001x}{0.001x + 0.000999} \\ &= \frac{0.001 \cdot 0.891}{0.001 \cdot 0.891 + 0.000999} \approx 0.47 \approx 0.5. \end{aligned}$$

(b) Using the definition of the conditional probability, we get that

unseen ↓

$$P\left(\max\{X, Y\} > \frac{3}{4} \middle| \min\{X, Y\} > \frac{1}{3}\right) = \frac{P\left(\max\{X, Y\} > \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right)}{P\left(\min\{X, Y\} > \frac{1}{3}\right)}.$$

We note that

1, A

$$\begin{aligned} P\left(\min\{X, Y\} > \frac{1}{3}\right) &= P\left(X > \frac{1}{3}, Y > \frac{1}{3}\right) = P\left(X > \frac{1}{3}\right) P\left(Y > \frac{1}{3}\right) \\ &= (1 - F(1/3))^2 = \left(\frac{2}{3}\right)^2 = \frac{4}{9}, \end{aligned}$$

where we used the fact that X and Y are independent and identically distributed with cdf $F(x) = x$ for $x \in [0, 1]$. Also, by the law of total probability, we have that

1, D

$$\begin{aligned} P\left(\max\{X, Y\} > \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right) &+ P\left(\max\{X, Y\} \leq \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right) \\ &= P\left(\min\{X, Y\} > \frac{1}{3}\right). \end{aligned}$$

Let us focus on the second term in the above equation:

1, D

$$\begin{aligned} P\left(\max\{X, Y\} \leq \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right) &= P\left(\frac{1}{3} < X \leq \frac{3}{4}, \frac{1}{3} < Y \leq \frac{3}{4}\right) \\ &= (F(3/4) - F(1/3))^2 = \left(\frac{3}{4} - \frac{1}{3}\right)^2 = \left(\frac{5}{12}\right)^2 = \frac{25}{144}. \end{aligned}$$

Altogether, we have

1, D

$$\begin{aligned} P\left(\max\{X, Y\} > \frac{3}{4} \middle| \min\{X, Y\} > \frac{1}{3}\right) &= \frac{P\left(\max\{X, Y\} > \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right)}{P\left(\min\{X, Y\} > \frac{1}{3}\right)} \\ &= \frac{P\left(\min\{X, Y\} > \frac{1}{3}\right) - P\left(\max\{X, Y\} \leq \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right)}{P\left(\min\{X, Y\} > \frac{1}{3}\right)} \\ &= 1 - \frac{P\left(\max\{X, Y\} \leq \frac{3}{4}, \min\{X, Y\} > \frac{1}{3}\right)}{P\left(\min\{X, Y\} > \frac{1}{3}\right)} \\ &= 1 - \frac{25 \cdot 9}{144 \cdot 4} = 1 - \frac{25}{16 \cdot 4} = 1 - \frac{25}{64} = \frac{39}{64}. \end{aligned}$$

2, D

(c) (i) We note that

$$\begin{aligned}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy &= \int_0^1 \int_0^y cx(y+1) dx dy = \int_0^1 c \frac{x^2}{2} (y+1) \Big|_{x=0}^y dy \\ &= \int_0^1 \frac{c}{2} (y^3 + y^2) dy = \frac{c}{2} \frac{1}{4} y^4 + \frac{c}{2} \frac{1}{3} y^3 \Big|_{y=0}^1 = c \left(\frac{1}{8} + \frac{1}{6} \right) = c \frac{7}{24} = 1 \iff c = \frac{24}{7}.\end{aligned}$$

Also, $f_{X,Y}$ is nonnegative for this choice of c .

1, A

(ii) We need to integrate out y . More precisely, we have, for $x \in (0, 1)$,

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_x^1 cx(y+1) dy = cx(0.5y^2 + y) \Big|_{y=x}^1 \\ &= cx(1.5 - 0.5x^2 - x) = \frac{24}{7}x \left(\frac{3}{2} - \frac{1}{2}x^2 - x \right) = \frac{12}{7}(-x^3 - 2x^2 + 3x),\end{aligned}$$

and $f_X(x) = 0$ otherwise.

2, A

(iii) We note that the conditional density of $Y|X = x$ is given by

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{cx(y+1)}{cx\left(\frac{3}{2} - \frac{1}{2}x^2 - x\right)} = \frac{y+1}{\frac{3}{2} - \frac{1}{2}x^2 - x},$$

for $0 < x < y < 1$. The conditional density and the corresponding conditional expectation is well-defined for all x such that $f_X(x) > 0$, i.e. for all $0 < x < y < 1$. Then,

1, B

$$\begin{aligned}E(Y|X=x) &= \int_{-\infty}^{\infty} y f_{Y|X=x}(y|x) dy = \int_x^1 \frac{y^2 + y}{\frac{3}{2} - \frac{1}{2}x^2 - x} dy \\ &= \frac{1}{\frac{3}{2} - \frac{1}{2}x^2 - x} \left(\frac{1}{3}y^3 + \frac{1}{2}y^2 \right) \Big|_{y=x}^1 \\ &= \frac{1}{\frac{3}{2} - \frac{1}{2}x^2 - x} \left(\frac{1}{3} + \frac{1}{2} - \frac{1}{3}x^3 - \frac{1}{2}x^2 \right) = \frac{\frac{5}{6} - \frac{1}{3}x^3 - \frac{1}{2}x^2}{\frac{3}{2} - \frac{1}{2}x^2 - x},\end{aligned}$$

which is well-defined for $0 < x < y < 1$.

2, B

4. (a) (i) The bias is computed using the linearity of expectation as:

sim. seen $\Downarrow$

$$b_{\theta}(\hat{\Theta}) = E[\hat{\Theta} - \theta] = E\left[\frac{1}{n+1} \sum_{i=1}^n X_i\right] - \theta = \frac{1}{n+1} \sum_{i=1}^n E[X_i] - \theta,$$

and so

1, A

$$\begin{aligned} b_{\theta}(\hat{\Theta}) &= \frac{1}{n+1} \sum_{i=1}^n \theta - \theta \\ &= \frac{n}{n+1} \theta - \theta \\ &= \frac{-1}{n+1} \theta. \end{aligned}$$

1, A

- (ii) Using Theorem 1.5.24, the mean squared error can be computed as

sim. seen $\Downarrow$

$$E[(\hat{\Theta} - \theta)^2] = [b_{\theta}(\hat{\Theta})]^2 + \text{Var}[\hat{\Theta}].$$

The bias has been computed in Part (i), and the variance is computed as

1, A

$$\begin{aligned} \text{Var}[\hat{\Theta}] &= \text{Var}\left[\frac{1}{n+1} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{(n+1)^2} \text{Var}\left[\sum_{i=1}^n X_i\right] \\ &= \frac{1}{(n+1)^2} \sum_{i=1}^n \text{Var}[X_i] \\ &= \frac{1}{(n+1)^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{n\sigma^2}{(n+1)^2} \end{aligned}$$

using the fact that the variables are independent in the third line. Therefore,

$$E[(\hat{\Theta} - \theta)^2] = \frac{\theta^2}{(n+1)^2} + \frac{n\sigma^2}{(n+1)^2} = \frac{\theta^2 + n\sigma^2}{(n+1)^2}.$$

2, A

Note that a direct solution without using this theorem is also possible. One method is to first use the fact that the X_i are i.i.d. with mean θ and variance σ^2 , and so using Proposition 1.2.6 in the notes $E[\bar{X}] = \theta$, and from Exercise 1.10 in the notes $E[\bar{X}^2] = \theta^2 + \frac{\sigma^2}{n}$. Then using the linearity of expectation

$$\begin{aligned}
E[(\hat{\Theta} - \theta)^2] &= E \left[\left(\frac{1}{n+1} \sum_{i=1}^n X_i - \theta \right)^2 \right] \\
&= E \left[\left(\frac{n}{n+1} \bar{X} - \theta \right)^2 \right] \\
&= E \left[\frac{n^2}{(n+1)^2} \bar{X}^2 - \frac{2n}{n+1} \bar{X} \theta + \theta^2 \right] \\
&= \frac{n^2}{(n+1)^2} E[\bar{X}^2] - \frac{2n\theta}{n+1} E[\bar{X}] + E[\theta^2] \\
&= \frac{n^2}{(n+1)^2} \left(\theta^2 + \frac{\sigma^2}{n} \right) - \frac{2n\theta^2}{n+1} + \theta^2 \\
&= \frac{n^2}{(n+1)^2} \left(\theta^2 + \frac{\sigma^2}{n} \right) - \frac{(2n - (n+1))\theta^2}{n+1} \\
&= \frac{n^2\theta^2}{(n+1)^2} + \frac{n\sigma^2}{(n+1)^2} - \frac{(n-1)\theta^2}{n+1} \\
&= \frac{n^2\theta^2}{(n+1)^2} + \frac{n\sigma^2}{(n+1)^2} - \frac{(n-1)(n+1)\theta^2}{(n+1)^2} \\
&= \frac{n^2\theta^2}{(n+1)^2} + \frac{n\sigma^2}{(n+1)^2} - \frac{(n^2-1)\theta^2}{(n+1)^2} \\
&= \frac{[n^2 - (n^2-1)]\theta^2}{(n+1)^2} + \frac{n\sigma^2}{(n+1)^2} \\
&= \frac{\theta^2 + n\sigma^2}{(n+1)^2},
\end{aligned}$$

as calculated above.

- (b) Let us label the clinical trial volunteers from 1 to n , and for $i \in \{1, 2, \dots, n\}$ let X_i be the random variable with value $x_i = 1$ if the vaccinated volunteer develops severe symptoms within three months, and $x_i = 0$ otherwise.

sim. seen $\Downarrow$

1, A

By this construction, $X_i \sim \text{Bern}(p)$ where p is the unknown proportion we wish to estimate. Then, $E[X_i] = p$ and $\text{Var}[X_i] = p(1 - p)$.

1, A

Using Proposition 1.2.6 in the notes, $E[\bar{X}] = p$ and $\text{Var}[\bar{X}] = p(1 - p)/n$.

1, B

Also, using Proposition 1.1.6 or Corollary 1.1.7 in the notes,

$$\text{Var}[X_i] = p(1 - p) \leq \frac{1}{4}, \text{ and } \text{Var}[\bar{X}] \leq \frac{1}{4n}.$$

1, B

If we estimate p using $\hat{p} = \bar{x}$, then for any $\epsilon > 0$, we can use Chebyshev's Inequality from the notes with $Y = \bar{X}$, and assuming the X_i are independent, to obtain

$$\begin{aligned} P(|Y - p| \geq \epsilon) &\leq \frac{\text{Var}[Y]}{\epsilon^2} \\ \Rightarrow P(|\bar{X} - p| \geq \epsilon) &\leq \frac{\text{Var}[\bar{X}]}{\epsilon^2} \leq \frac{1}{4n\epsilon^2} \\ \Rightarrow P(|\bar{X} - p| < \epsilon) &\geq 1 - \frac{1}{4n\epsilon^2}. \end{aligned}$$

If we wish to estimate p to within $\delta = 0.02$ with confidence 0.95, setting $\epsilon = \delta = 0.02$, this implies we require

$$\begin{aligned} 1 - \frac{1}{4n\epsilon^2} &\geq 0.95 \\ \Rightarrow 0.05 &\geq \frac{1}{4n\epsilon^2} \\ \Rightarrow n &\geq \frac{1}{4\epsilon^2(0.05)} = \frac{1}{4\delta^2(0.05)} = \frac{1}{4(0.02)^2(0.05)} = 12500. \end{aligned}$$

2, B

There is perhaps a slight ambiguity in the phrase "within $\delta = 0.02$ ", so if $\epsilon = 0.5\delta = 0.01 \Rightarrow n \geq 50000$ is calculated, this can also be accepted.

unseen $\Downarrow$

- (c) The median m for any random variable Y is by definition the value such that $P(Y \geq m) \geq 0.5$ and $P(Y \leq m) \geq 0.5$.

If Y is continuous, then $P(Y > m) = P(Y \geq m) \geq 0.5$, and since necessarily $P(Y \leq m) + P(Y > m) = P(Y \in \mathbb{R}) = 1$ this implies $P(Y \leq m) = 0.5$

(since the above implies $P(Y \leq m) = 1 - P(Y > m) \leq 0.5$).

1, D

We first note that since $\Gamma(n) = (n - 1)!$, when n is a positive integer, and so $\Gamma(1) = (0)! = 1$, and

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha) \cdot 1} = \alpha,$$

using the hint provided. Then

$$\begin{aligned} 0.5 = P(Y \leq m) &= \int_0^m \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1 - y)^{\beta-1} dy = \int_0^m \alpha y^{\alpha-1} dy = [y^\alpha]_0^m = m^\alpha \\ \Rightarrow \alpha \log m &= \log 0.5 = -\log 2 \\ \Rightarrow \log m &= -\frac{1}{\alpha} \log 2 = \log 2^{-1/\alpha} \\ \Rightarrow m &= 2^{-1/\alpha}, \end{aligned}$$

as required.

2, C

- (d) (i) Since the true variance is unknown, but the sample variance is provided, we can use Student's t -distribution. Defining

sim. seen ↓

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

where $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$, then $T \sim t_{n-1}$, Student's t distribution with $n - 1$ degrees of freedom.

1, A

Then, for any $\alpha \in (0, 1)$, $t_{n-1, 1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile, and by symmetry of the t -distribution, $t_{n-1, \alpha/2} = -t_{n-1, 1-\alpha/2}$. Then

$$\begin{aligned} P(t_{n-1, \alpha/2} \leq \frac{\mu - \bar{Z}}{S/\sqrt{n}} \leq t_{n-1, 1-\alpha/2}) &= 1 - \alpha \\ \Rightarrow P(t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \leq \mu - \bar{Z} \leq t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}) &= 1 - \alpha \\ \Rightarrow P(\bar{Z} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{Z} + t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}) &= 1 - \alpha \\ \Rightarrow P(\bar{Z} - t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{Z} + t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}) &= 1 - \alpha. \end{aligned}$$

1, A

Using $n = 20$ and $\alpha = 0.01$, we can use Table 2 to find $t_{19, 0.995} = 2.861$.

Then the realised confidence interval, using $\bar{z} = 10.5$ and $s^2 = 580$ becomes

1, A

$$\left(10.5 - 2.861 \frac{\sqrt{580}}{\sqrt{20}}, 10.5 + 2.861 \frac{\sqrt{580}}{\sqrt{20}} \right) = (-4.91, 25.91).$$

1, A

sim. seen ↓

- (ii) Since the confidence interval $(-4.91, 25.91)$ in Part (i) contains 0, there is no evidence that the daily changes in the share price over the past n days have generally been increases (or, in other words, that the share price has generally been increasing).

2, D

As an aside, note that a calculator is not necessary for this conclusion, because one could approximate

$$10.5 - 2.861 \frac{\sqrt{580}}{\sqrt{20}} = 10.5 - 2.861 \sqrt{29} < 10.5 - 2.5 \sqrt{25} = 10.5 - 12.5 = -2,$$

and similarly $10.5 + 2.861 \frac{\sqrt{580}}{\sqrt{20}} > 23$, and the approximated (sub)interval $(-2, 23)$ also contains 0.

5. (a) (i) The Type I error $\alpha = P(\text{Reject } H_0 | H_0 \text{ is true})$.
The Type II error $\beta = P(\text{Fail to reject } H_0 | H_0 \text{ is false})$.
- (ii) First, statistical power is defined as $1 - \beta$, where β is the Type II error, so these two quantities are always closely related.
Second, we cannot maximise statistical power without affecting the Type I error α . In fact, while we can control the level of α for our hypothesis test, we cannot set β (and hence cannot set $1 - \beta$), since this is a consequence of our test at the specified level α . Therefore, in order to maximise statistical power, we need to adjust the value of α .
- (iii) If H_0 is that the defendant is innocent, then the Type I error is the probability of convicting the defendant, when the defendant is innocent.
The Type II error is the probability of acquitting the defendant when the defendant is in fact guilty.

seen ↓

2, A

1, B

1, C

unseen ↓

1, D

1, D

(b) (i) The null hypothesis is

sim. seen ↓

$$H_0 : \theta \leq 185.$$

The alternative hypothesis is

$$H_1 : \theta > 185.$$

2, B

(ii) Defining the statistic

$$Z = \frac{\theta_0 - \bar{X}}{\sigma/\sqrt{n}},$$

where $\theta_0 = 185$, $\sigma = 3$ and $n = 15$, then Z follows a normal distribution with mean $\frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) \geq 0$ and variance 1. Then we can write

1, B

$$\begin{aligned} P(Z \geq z_\alpha) &= 1 - \alpha \\ \Rightarrow P\left(\frac{\theta_0 - \bar{X}}{\sigma/\sqrt{n}} \geq z_\alpha\right) &= 1 - \alpha \\ \Rightarrow P(\theta_0 < \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}) &= \alpha, \end{aligned}$$

where z_α is the α -quantile for the standard normal distribution. In this case, $\alpha = 0.05$ and $z_{0.05} = -1.645$ (interpolating between 1.64 and 1.65 in Table 1 for 0.95 and using symmetry, although either of these two values would also be acceptable here). Since $\bar{x} = 187$,

1, B

1, B

$$185 = \theta_0 < 187 - 1.645 \frac{3}{\sqrt{15}} \approx 187 - 1.27 = 185.73,$$

and so the null hypothesis is rejected at significance level $\alpha = 0.05$, i.e. $\theta \not\leq 185$ at significance level $\alpha = 0.05$.

1, B

An alternative approach would be to simply use

$$P(Z < z_\alpha) = \alpha,$$

so if $z = \frac{\theta_0 - \bar{x}}{\sigma/\sqrt{n}} < z_\alpha$, then the null hypothesis would be rejected. One computes

$$z = \frac{\theta_0 - \bar{x}}{\sigma/\sqrt{n}} = \frac{185 - 187}{3/\sqrt{15}} \approx -2.582 < -1.645 = z_\alpha,$$

so the null hypothesis is rejected.

- (c) (i) Although $5 \times 10^{-6} < 0.001$ and $2 \times 10^{-5} < 0.001$, we need to correct for the 1000 hypotheses that were tested (the different genetic mutations).
Using the Bonferroni correction, we need to use the adjusted threshold

sim. seen ↓

$$\frac{\alpha}{n} = \frac{0.001}{1000} = 1 \times 10^{-6}.$$

1, C

Then, since $5 \times 10^{-6} \not< 1 \times 10^{-6}$ and $2 \times 10^{-5} \not< 1 \times 10^{-6}$, neither of these two p -values would be declared significant.

1, C

- (ii) We would still need to correct for multiple testing, but we could perhaps make the original threshold α less conservative.

We need to choose a new threshold α' so that $5 \times 10^{-6} < \alpha'/n < 2 \times 10^{-5}$, and then exactly one p -value would be significant; for example, the threshold $\alpha'/n = 1 \times 10^{-5}$ would work. Then

$$\begin{aligned} \frac{\alpha'}{n} &= 1 \times 10^{-5} \\ \Rightarrow \alpha' &= n \cdot 1 \times 10^{-5} = 1000 \cdot 1 \times 10^{-5} = 0.01. \end{aligned}$$

2, D

- (d) The sample correlation is defined as

unseen ↓

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

From an exercise in the notes, we have

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = n \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right].$$

Then we can rewrite r_{XY} as

1, B

$$\begin{aligned} r_{XY} &= \frac{\sum_{i=1}^n (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x} \bar{y})}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}} \\ &= \frac{\sum_{i=1}^n x_i y_i - \bar{x} \sum_{i=1}^n y_i - \bar{y} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x} \bar{y}}{\sqrt{n \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right]} \sqrt{n \left[\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right]}} \\ &= \frac{\sum_{i=1}^n x_i y_i - n(\bar{x})(\bar{y}) - n(\bar{y})(\bar{x}) + n(\bar{x})(\bar{y})}{n \sqrt{\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right]} \sqrt{\left[\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right]}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - (\bar{x})(\bar{y})}{\sqrt{\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right]} \sqrt{\left[\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right]}}. \end{aligned}$$

Substituting in the given values,

2, C

$$r_{XY} = \frac{(37.76) - (5.34)(4.72)}{\sqrt{[38.86 - (5.34)^2]} \sqrt{[(50.92) - (4.72)^2]}} = 0.7212776.$$

1, A

6. (a) Given that the independent random variables $X_1, X_2, \dots, X_n$ are observed as $x_1, x_2, \dots, x_n$, the joint likelihood is computed as

sim. seen ↓

$$\begin{aligned} L(\theta|\mathbf{x}) &= f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) \\ &= \prod_{i=1}^n \frac{1}{x_i \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\log x_i - \theta)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \frac{1}{\prod_{i=1}^n x_i} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \theta)^2\right). \end{aligned}$$

An equivalent likelihood is

$$L(\theta|\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \theta)^2\right),$$

and the log-likelihood is

$$\log L(\theta|\mathbf{x}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \theta)^2.$$

1, A

Maximising the log-likelihood is equivalent to maximising the likelihood, and to find the maximum of this function we start by computing the first derivative

$$\frac{d}{d\theta} \log L(\theta|\mathbf{x}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(\log x_i - \theta)(-1) = \frac{1}{\sigma^2} \sum_{i=1}^n (\log x_i - \theta).$$

Setting the first derivative equal to zero,

1, A

$$\begin{aligned} \frac{d}{d\theta} \log L(\theta|\mathbf{x}) &= 0 \\ \Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n (\log x_i - \theta) &= 0 \\ \Rightarrow \sum_{i=1}^n (\log x_i - \theta) &= 0 \\ \Rightarrow \sum_{i=1}^n \log x_i - n\theta &= 0 \\ \Rightarrow \theta &= \frac{1}{n} \sum_{i=1}^n \log x_i = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right). \end{aligned}$$

1, A

To check this is a maximum, we first need to calculate the second derivative

$$\frac{d^2}{d\theta^2} \log L(\theta|\mathbf{x}) = \frac{d}{d\theta} \left(\frac{d}{d\theta} \log L(\theta|\mathbf{x}) \right) = \frac{d}{d\theta} \left(\frac{1}{\sigma^2} \log \left(\prod_{i=1}^n x_i \right) - \frac{n}{\sigma^2} \theta \right) = -\frac{n}{\sigma^2} < 0,$$

which shows that $\theta = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right)$ is a local maximum.

1, A

To check it is a global maximum, we need to check the limits:

$$\lim_{\theta \rightarrow -\infty} \log L(\theta|\mathbf{x}) = \lim_{\theta \rightarrow -\infty} \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \theta)^2 \right] = -\infty.$$

Similarly, $\lim_{\theta \rightarrow \infty} \log L(\theta|\mathbf{x}) = -\infty$.

1, A

Therefore, $\theta = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right)$ is a global maximum, and $\hat{\theta} = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right)$ is the maximum likelihood estimate of θ .

Finally, this means that $\hat{\Theta} = \frac{1}{n} \log \left(\prod_{i=1}^n X_i \right)$ is the maximum likelihood estimator of θ .

1, A

An alternative approach to computing the maximum likelihood estimate of θ is to set $z_i = \log x_i$ for $i \in \{1, 2, \dots, n\}$, and then using an exercise from the notes,

$$\begin{aligned} \sum_{i=1}^n (z_i - \theta)^2 &\geq \sum_{i=1}^n (z_i - \bar{z})^2 \\ \Rightarrow -\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \theta)^2 &\leq -\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \bar{z})^2 \\ \Rightarrow \log L(\theta|\mathbf{x}) &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \theta)^2 \leq -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \bar{z})^2, \end{aligned}$$

where

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i = \frac{1}{n} \sum_{i=1}^n \log x_i = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right),$$

which shows that $\hat{\theta} = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right)$ is the maximum likelihood estimate.

sim. seen ↓

- (b) (i) Yes, this is a bootstrap sample because it also has five elements (the same number as in the original sample), and each of the elements is an element in the original sample.

1, A

- (ii) No, this is not a bootstrap sample because 76 is not an element in the original sample.

1, A

sim. seen ↓

- (c) The model does not fit the data well, because the residual plot shows that the residual values $\hat{\epsilon}_i$ are not independent (and the model assumes they are independent).

2, C

- (d) In the notes we derived that the maximum likelihood estimates for the simple linear regression model parameters are

unseen ↓

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

Then the fitted model becomes

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\epsilon}_i,$$

where the $\hat{\epsilon}_i$ are the residuals, and the fitted regression line is

$$y = \hat{\beta}_0 + \hat{\beta}_1 x.$$

1, D

Since

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

this shows that $(\bar{x}, \bar{y})$ should be a point on the fitted regression line, if the maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ have been calculated correctly.

2, D

Therefore, the linear regression line that has been plotted is incorrect.

1, D

sim. seen ↓

- (e) The probability density function for each X_i is given as

$$f(x_i|\theta) = \mu\theta \exp(\theta + \mu x_i - \theta e^{\mu x_i}),$$

where $x_i \geq 0$. Since the X_i are assumed to be independent, then

1, A

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f(x_i|\theta) \\ &= \prod_{i=1}^n \mu\theta \exp(\theta + \mu x_i - \theta e^{\mu x_i}) \\ &= \mu^n \theta^n \exp\left(\sum_{i=1}^n [\theta + \mu x_i - \theta e^{\mu x_i}]\right) \\ &= \mu^n \theta^n \exp\left(n\theta + \mu n\bar{x} - \theta \sum_{i=1}^n e^{\mu x_i}\right) \\ &= \mu^n \theta^n \exp\left(\theta \left[n - \sum_{i=1}^n e^{\mu x_i}\right] + \mu n\bar{x}\right). \end{aligned}$$

1, A

The prior distribution of $\theta \in (0, \infty)$ is given as

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}.$$

The posterior probability density function can then be computed as

$$\begin{aligned}
 \pi(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\
 &= \mu^n \theta^n \exp\left(\theta\left[n - \sum_{i=1}^n e^{\mu x_i}\right] + \mu n \bar{x}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\
 &\propto \theta^n \exp\left(\theta\left[n - \sum_{i=1}^n e^{\mu x_i}\right] + \mu n \bar{x}\right) \theta^{\alpha-1} e^{-\beta\theta} \\
 &= \theta^{n+\alpha-1} \exp\left(\theta\left[n - \sum_{i=1}^n e^{\mu x_i}\right] + \mu n \bar{x} - \beta\theta\right) \\
 &= \theta^{n+\alpha-1} \exp\left(-\theta\left[\beta + \sum_{i=1}^n e^{\mu x_i} - n\right]\right) \exp(\mu n \bar{x}) \\
 &\propto \theta^{n+\alpha-1} \exp\left(-\left[\beta + \sum_{i=1}^n e^{\mu x_i} - n\right]\theta\right).
 \end{aligned}$$

3, B

Therefore we see that the posterior distribution, given an appropriate scaling constant, is a $\text{Gamma}(n + \alpha, \beta + \sum_{i=1}^n e^{\mu x_i} - n)$ distribution with probability density function

$$\pi(\theta|\mathbf{x}) = \frac{(\beta + \sum_{i=1}^n e^{\mu x_i} - n)^{n+\alpha}}{\Gamma(n + \alpha)} \theta^{n+\alpha-1} \exp\left(-\left[\beta + \sum_{i=1}^n e^{\mu x_i} - n\right]\theta\right).$$

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

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	Question	Comments for Students
		(ai) Common mistake to use the generic example $\{\text{emptyset}, A, A^c, \Omega\}$ without explicitly defining $A^c = \{B, C, D\}$ or $\Omega = \{A, B, C, D\}$ to link this to the question. There was a variety of incorrect notation for sets of cardinality greater than one, e.g. referring $\{A, B\}$ as $\{A \cup B\}$. (aii) Generally well done. Occasionally students would consider $\text{Im}(X)$ but the the pre-image. (aiii) Often students carried on with X from part aii without linking it to Y in this question, losing marks. Many also did not define the probability measure they were using. (bi) If attempted, this was generally well done with students taking a mix of induction and direct manipulation approaches. (biii) Not attempted by many students. A common error here was to take a different "story" for
Probability and Statistics_MATH40005	1	each side of the equality and not draw these together to make the equivalence plain.

There were very few totally disastrous scripts but at the same time there was evidence that several students had a limited understanding of what is meant by a proof in (a) as they started by assuming that covariances are bilinear. Part (b) (i) was straightforward and done well. In part (b) (ii) those who did not use part (a) often ended up with some unnecessarily complicated expressions and inevitable slips. Some of these slips were arithmetic but others showed a basic lack of understanding about how to manipulate variances and covariances.

Part (c) separated those who had a good understanding of the difference between continuous and discrete random variables from those who guessed what was going on. There were some very good answers but also many who failed to notice that one needed to use the law of Total Probability to find the c.d.f of Z.

Part a. was looking for an implementation of the Bayes Theorem and the Law of Total Probability for checking the journalist claims. Alternative solutions were taken as correct, whether you have attempted to answer the question by discussing the assumption made $P(\text{Positive test}|\text{covid})=1$ or by attempting to compute this probability using the Bayes Theorem.

Part b. question was looking for an implementation of the conditional probability. Some very good answers were given, more justifications/details about the computations were needed for gaining full marks.

Part c. was overall found the least challenging question. Some of you lost marks as the wrong integration limits were implemented.

Two common issues were Q4(a)(ii): (1) not using/quoting the result linking the mean squared error of an estimator with its bias and variance; (2) Q4(b) - many students did not use Chebyshev's inequality.

"This question was addressed reasonably well by most candidates.

5 Part (a)(i) was answered well by nearly all candidates.

(a) Common error was to consider only one observation rather than a vector of observed values. Problems also arose from people working with likelihood directly, rather than log-likelihood.

(b) Answered uniformly very well by the majority of those who answered at all.

(c) Many responses commenting on non-normality rather than lack of independence in the residuals. There is insufficient evidence from these plots to make such distributional claims.

(d) Common mistake to say that the line had about equal number of points above and below so was fine. Another common mistake was to say that $y = ax + b + \text{error}$ so $y(\bar{x}) = \bar{y} + \text{error}$, or similar, to justify the line not intersecting this point.

6 (e) Done very well by many. Quite a few students mis-identified the gamma as a beta distribution.