Imperial College London

MATH40005

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2020

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

Probability and Statistics

Date: 14th May 2020

Time: 09.00am – 12.00 noon (BST)

Time Allowed: 3 Hours

Upload Time Allowed: 30 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.

© 2020 Imperial College London

Throughout the exam, we assume that $(\Omega, \mathcal{F}, \mathrm{P})$ 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) Define a σ-algebra. (3 marks)

   (b) Define a probability measure on $(\Omega, \mathcal{F})$. (3 marks)

   (c) A diagnostic test has a probability 0.9 of giving a positive result when applied to a person suffering from a certain disease, and a probability 0.2 of giving a (false) positive when applied to a non-sufferer. It is estimated that 10 % of the population are sufferers. Suppose that the test is now administered to a person about whom we have no relevant information relating to the disease (apart from the fact that he/she comes from this population). Calculate the following probabilities:

      (i) that the test result will be positive; (3 marks)

      (ii) that, given a positive result, the person is a sufferer; (3 marks)

   (d) How many possibilities are there to write the number 7 as an ordered sum of 3 positive integers? [E.g. 7=1+3+3 would be one possible case and 7=3+1+3 would be another case.] (4 marks)

   (e) Let $k, n \in \mathbb{N} = \{1, 2, \ldots\}$. How many possibilities are there to write the number k as an ordered sum of n positive integers? (4 marks)

(Total: 20 marks)

2. (a) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space with $\Omega = \{1, 2, \ldots, 10\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ (the power σ-algebra of Ω). Let $X : \Omega \to \mathbb{R}$ with $X(\omega) = \omega + 5$. Prove that X is a discrete random variable. (3 marks)

(b) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space with $\Omega = \{1, 2, \ldots, 100\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ (the power σ-algebra of Ω). Consider a discrete random variable X on this probability space with probability mass function given by $\mathrm{P}(X = 3) = \frac{1}{2}, \mathrm{P}(X = 5) = \frac{1}{5}, \mathrm{P}(X = 100) = \frac{3}{10}, \mathrm{P}(X = x) = 0$ for $x \notin \{3, 5, 100\}$. Find the cumulative distribution function F_X of X. (4 marks)

(c) Let X and Y denote independent geometric random variables with parameters p_1 and p_2, respectively, where $p_1, p_2 \in (0, 1)$. (Please refer to the hint below for the probability mass function of a geometric random variable.)

(i) Derive the cumulative distribution functions of X and Y. (3 marks)

(ii) Show that $Z = \min\{X, Y\}$ follows a geometric distribution and find the corresponding parameter. (4 marks)

(d) Imagine you toss a fair coin repeatedly. You denote by H the outcome Heads and by T the outcome Tails. How many times, on average, do you need to toss the coin to see the pattern HT (i.e. Heads followed by Tails) for the first time? (6 marks)

(Total: 20 marks)

Hint: If X is **geometrically** distributed with parameter $p \in (0, 1)$, then its probability mass function p_X is given by

$$p_X(x) = \begin{cases} (1-p)^{x-1}p, & \text{for } x = 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

3. (a) Which properties does a function $f : \mathbb{R} \to \mathbb{R}$ need to satisfy in order to be a valid probability density function? (2 marks)

(b) For each of the functions $f(x)$ given below determine whether $f(x)$ is a valid probability density function (p.d.f.). If $f(x)$ is not a valid p.d.f., determine if there exists a constant c such that $cf(x)$ is a valid p.d.f.. Note that in each case, $f(x) = 0$ for all x not in the interval(s) specified.

(i) $f(x) = 3x$ for $0 < x < 1$, (2 marks)

(ii) $f(x) = -1$ for $0 < x < 1$, (2 marks)

(iii) $f(x) = 1$ for $0 < x < 1$ and $f(x) = -1$ for $1 < x < 2$. (2 marks)

(c) Consider three jointly continuous random variables X, Y, Z with joint probability density function given by

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

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

(i) Show that $c = 6$. (3 marks)

(ii) Find $\mathrm{E}(XYZ)$. (3 marks)

(d) (i) Define a partition of the sample space Ω and give an example of $(\Omega, \mathcal{F})$ with a valid partition. (3 marks)

(ii) Prove the law of the total expectation for a discrete random variable X. I.e. consider a partition $\{B_i : i \in \mathcal{I}\}$ of Ω with $\mathrm{P}(B_i) > 0$ for all $i \in \mathcal{I}$. Let X denote a discrete random variable with finite expectation. Show that

$$\mathrm{E}(X) = \sum_{i \in \mathcal{I}} \mathrm{E}(X|B_i)\mathrm{P}(B_i),$$

whenever the sum converges absolutely. (3 marks)

(Total: 20 marks)

4. (a) Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and each follows a normal distribution with mean μ and variance σ^2. We define the following estimators

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad Z = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

where $\overline{X}$ is an estimator of μ, and S^2 and Z are estimators of σ^2. Carefully justifying all your steps and stating any results used:

(i) Find the distribution of $\overline{X}$. (3 marks)

(ii) Given that $\mathrm{E}(S^2) = \sigma^2$, compute $\mathrm{E}(Z)$. (1 mark)

(iii) Given that $\mathrm{Var}(S^2) = \dfrac{2\sigma^4}{n-1}$, compute $\mathrm{Var}(Z)$. (1 mark)

(iv) Compute the bias of Z. (1 mark)

(v) Compute the mean squared error of Z. (2 marks)

(vi) Choose a constant b so that the quantity bZ has a chi-squared distribution, and state the degrees of freedom for this chi-squared distribution. (2 marks)

(vii) Compute $\mathrm{Cov}(\overline{X}, Z)$. (2 marks)

(b) Markov's inequality states that if a random variable X can only take nonnegative values, then

$$\mathrm{P}(X \geq a) \leq \frac{\mathrm{E}(X)}{a}, \qquad \text{for all } a > 0.$$

Prove Markov's inequality. (4 marks)

(c) Suppose that a medical research lab is testing for the association of different genetic variants with a particular disease. The research team decides in advance that a significance threshold of $\alpha = 0.01$ will be used for each test. A total of 100 genetic variants are tested for association based on the data the team has available. The following table lists the five smallest p-values (in decreasing order) and the genetic variants for which these p-values were found:

Genetic variant	A	B	C	D	E
p-value from test	3×10^{-2}	9×10^{-3}	4×10^{-4}	2×10^{-5}	5×10^{-6}

Which of the genetic variants in the table (if any) should the research team declare to be significantly associated with the disease given the data, the statistical test and the significance threshold that were used? Provide justification and state any results used. (4 marks)

(Total: 20 marks)

5. (a) Given a sample of real-valued observations $x_1, x_2, \ldots, x_n$, prove that for any constant $a \in \mathbb{R}$

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 \le \sum_{i=1}^{n}(x_i - a)^2,$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of the observations. (4 marks)

(b) Given n pairs of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ for quantities X and Y, define the sample means $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and define

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

Consider the model given by

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i \in \{1, 2, \ldots, n\},$$

where the e_i, $i \in \{1, 2, \ldots, n\}$, are unobservable errors. Find the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the parameters β_0 and β_1, respectively, such that

$$\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 = \min_{b_0, b_1} \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2.$$

(7 marks)

(c) Suppose we have two samples of data, independent observations $x_1, x_2, \ldots, x_n$ of the random variable X and independent observations $y_1, y_2, \ldots, y_m$ of the random variable Y. We wish to use the two-sample t-test to decide whether or not $\mu_X = \mathrm{E}(X)$ and $\mu_Y = \mathrm{E}(Y)$ are equal.

(i) What is the null hypothesis for the t-test in this case? (1 mark)

(ii) What assumptions are required in order to have theoretical justification for conducting the t-test in this case? (2 marks)

(d) Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed according to a uniform distribution on the closed interval $[0, \theta]$, for some parameter $\theta > 0$, where the exact value of the parameter θ is unknown. Given that $\mathbf{X} = (X_1, X_2, \ldots X_n)$ is observed as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, find the maximum likelihood estimator of θ. Provide justification for all of your steps. (6 marks)

(Total: 20 marks)

Hint: If X is **uniformly** distributed on the interval $[a, b]$, with $a < b$, then its probability density function f_X is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } x \in [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

Note that this question is split over two pages. Please turn the page to see the rest of Question 6.

6. (a) Suppose that the random variables $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed according to a distribution F_Y, which has an unknown mean μ that we wish to estimate. Suppose that $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ is observed as $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, and we are given that

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 5, \qquad \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = 4, \qquad n = 10.$$

Noting that you have access to Tables 1 and 2 below:

(i) If we can assume that the random variables $Y_1, Y_2, \ldots, Y_n$ are normally distributed with variance $\mathrm{Var}(Y) = \sigma^2 = 9$, construct a 90% confidence interval for the unknown mean μ based on the data $\mathbf{y}$. (2 marks)

(ii) If we can assume that the random variables $Y_1, Y_2, \ldots, Y_n$ are normally distributed but the variance $\mathrm{Var}(Y) = \sigma^2$ is unknown, construct a 95% confidence interval for the unknown mean μ based on the data $\mathbf{y}$. (2 marks)

(iii) If we cannot assume that the random variables $Y_1, Y_2, \ldots, Y_n$ are normally distributed but we can assume that the variance is $\mathrm{Var}(Y) = \sigma^2 = 16$, construct a confidence interval for the unknown mean $\mu = \mathrm{E}(Y)$ which has coverage probability at least 0.99, whatever the distribution of F_Y. (2 marks)

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

ν	0.90	0.95	0.975	0.99
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764

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

z	$\mathrm{P}(Z < z)$
1.281	0.900
1.645	0.950
1.960	0.975
2.326	0.990

[IMPORTANT: Question 6 continues on the next page.]

[**Question 6 continues on this page**]

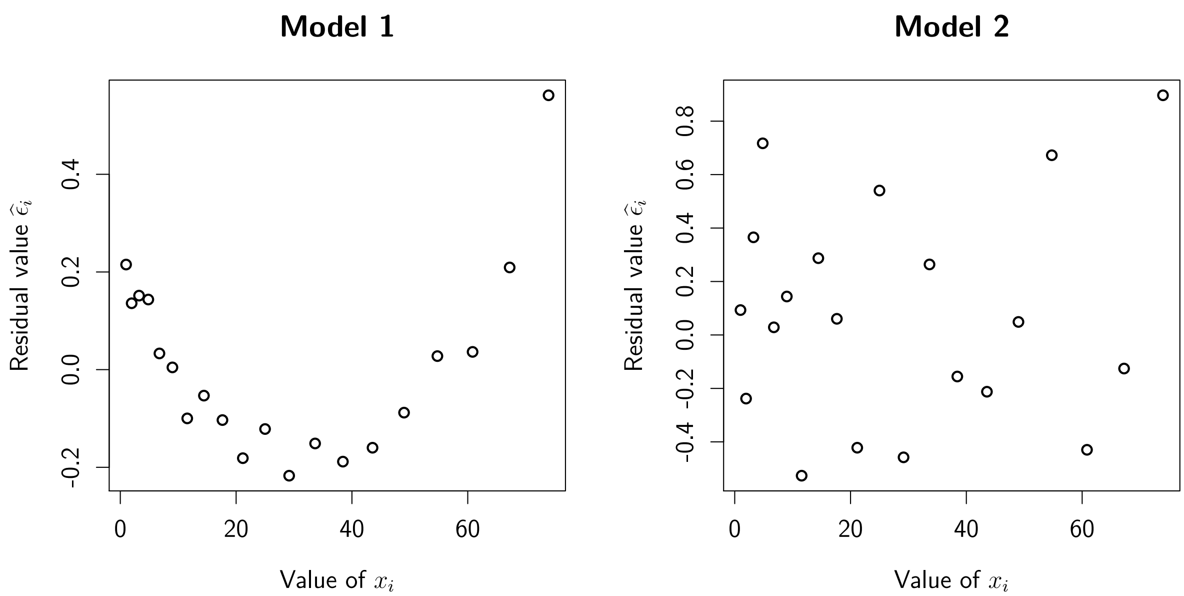
(b) Suppose one fits a simple linear regression model to the data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ as

$$Y_i = \beta_0 + \beta_1 g(x_i) + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

where $g : \mathbb{R} \to \mathbb{R}$ is some univariate transformation and $n = 20$.

(i) What joint distribution are the errors ϵ_i assumed to follow? (1 mark)

(ii) For two different choices of transformation g, one has two models with the fitted residuals $\widehat{\epsilon}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 g(x_i)$ shown in the figures below. For each model, state whether the model fits the data well or not and justify your answer. (4 marks)



(c) Suppose that the random variables $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed as an exponential distribution with unknown parameter θ, which has probability density function

$$f(z) = \theta \exp(-\theta z), \qquad \text{with support } z > 0.$$

Following a Bayesian approach and assuming that θ is a random variable with a $\Gamma(\alpha, \beta)$ prior which has probability density function

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta), \qquad \text{with support } \theta > 0,$$

and given that $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ is observed as $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, find the posterior distribution of θ given $\mathbf{z}$ and give the name of this distribution. (3 marks)

(d) Suppose that a random variable X has mean $\mathrm{E}(X) = 2$, another random variable Y has mean $\mathrm{E}(Y) = 3$, and it is known that $\mathrm{E}(XY) = 4$. It is also known that $2 \le Y \le 5$. Find a nontrivial lower bound on the standard deviation of X. (6 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2020

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
. . . .	. . . .	. . . .

© 2020 Imperial College London

1. (a) Let Ω be a set. A collection of subsets of Ω denoted by $\mathcal{F}$ is called a *σ-algebra* if seen $\Downarrow$

   1. $\emptyset \in \mathcal{F}$,
   2. $\mathcal{F}$ is closed under complements, i.e. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$, and
   3. $\mathcal{F}$ is closed under countable union, i.e. $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$. 3, A

(b) A mapping $\mathrm{P} : \mathcal{F} \to \mathbb{R}$ is called a *probability measure* on $(\Omega, \mathcal{F})$ if it satisfies three conditions: seen $\Downarrow$

   (i) $\mathrm{P}(A) \geq 0$ for all events $A \in \mathcal{F}$,

   (ii) $\mathrm{P}(\Omega) = 1$,

   (iii) For any sequence of disjoint events $A_1, A_2, A_3, \dots \in \mathcal{F}$ we have

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i).$$

   [Note that by "disjoint events" we mean that $A_i \cap A_j = \emptyset$ for all $i \neq j$.] 3, A

(c) Let $T \equiv$ "Test positive", $S \equiv$ "Sufferer". Then $\mathrm{P}(T|S) = 0.9$, $\mathrm{P}(T|S^c) = 0.2$, $\mathrm{P}(S) = 0.1$. sim. seen $\Downarrow$

   (i) Using the law of total probability, we have

$$\mathrm{P}(T) = \mathrm{P}(T|S)\mathrm{P}(S) + \mathrm{P}(T|S^c)\mathrm{P}(S^c) = \frac{9}{10} \times \frac{1}{10} + \frac{2}{10} \times \frac{9}{10} = 0.27.$$

3, B

   (ii) By Bayes' formula, we have

$$\mathrm{P}(S|T) = \frac{\mathrm{P}(T|S)\mathrm{P}(S)}{\mathrm{P}(T)} = \frac{\frac{9}{10} \times \frac{1}{10}}{\frac{27}{100}} = \frac{1}{3}.$$

3, B

(d) We present an elementary solution here, followed by the general solution in Part (b): We can write down all possible 3-tuples of numbers which sum up to 7: unseen $\Downarrow$

   * $(1, 3, 3)$ with $3!/2! = 3$ possible arrangements (where we adjusted for over-counting since the number 3 appears twice),
   * $(1, 1, 5)$ with $3!/2! = 3$ possible arrangements,
   * $(2, 2, 3)$ with $3!/2! = 3$ possible arrangements,
   * $(1, 2, 4)$ with $3! = 6$ possible arrangements.

   Hence there are $3 + 3 + 3 + 6 = 15$ possibilities of writing the number 7 as a sum of 3 positive integers. 4, B

(e) We can use a stars and bars argument: unseen $\Downarrow$

1, C

   Clearly, there are 0 possibilities if $k < n$. Suppose that $k \geq n$. Then we represent the number k as k stars which we would like to place in n bins such that each bin contains at least one object (since we have the restriction that all addends are *positive* integers). We can first write the k stars in one line. Then there are $k - 1$ possible gaps between the stars, where a bar could be inserted to separate the bins. We need to select $n - 1$ gaps out of the $k - 1$ gaps, to create the n bins, so in total we have $\binom{k-1}{n-1}$ possibilities. 3, D

   E.g. in the case of Part (a), one possible configuration is $*| * * * | * **$, where we have 1 star in the first bin, followed by 3 stars in the next bin, followed by 3 stars in the last bin.

2. (a) We check the two defining properties of a discrete random variable: $\boxed{\text{meth seen} \Downarrow}$

   * ImX is a countable subset of $\mathbb{R}$ since Im$X = \{\omega + 5 : \omega \in \Omega\} = \{6, 7, \dots, 15\}$ is a finite set.
   * $X^{-1}(x) \in \mathcal{F}$ for all $x \in \mathbb{R}$ since, $\boxed{1, \text{C}}$
     · for all $x \in$ ImX, we have $X^{-1}(x) = \{\omega \in \Omega : X(\omega) = x\} = \{x - 5\} \in \Omega$ and since $\mathcal{F} = \mathcal{P}(\Omega)$, we have that $X^{-1}(x) \in \mathcal{F}$, and
     · for all $x \in \mathbb{R} \setminus$ ImX, we have $X^{-1}(x) = \emptyset \in \mathcal{F}$. $\boxed{2, \text{C}}$

(b) The cumulative distribution function is given by $\boxed{\text{meth seen} \Downarrow}$

$$F_X(x) = \mathrm{P}(X \leq x) = \begin{cases} 0, & \text{for } x \in (-\infty, 3), \\ \frac{1}{2}, & \text{for } x \in [3, 5), \\ \frac{7}{10}, & \text{for } x \in [5, 100), \\ 1, & \text{for } x \in [100, \infty). \end{cases}$$

$\boxed{4, \text{A}}$

(c) (i) First we derive the cumulative distribution function of a geometric distribution. Let $k \in \mathbb{N}$, then $\boxed{\text{meth seen} \Downarrow}$

$$F_X(k) = \mathrm{P}(X \leq k) = \sum_{x=1}^{k} (1 - p_1)^{x-1} p_1 = \sum_{x=0}^{k-1} (1 - p_1)^x p_1$$
$$\overset{\text{geom. series}}{=} p_1 \frac{1 - (1 - p_1)^k}{1 - (1 - p_1)} = 1 - (1 - p_1)^k.$$

So, in general we have $\boxed{2, \text{A}}$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1, \\ 1 - (1 - p_1)^{\lfloor x \rfloor}, & \text{if } x \geq 1. \end{cases}$$

Hence $\mathrm{P}(X > x) = (1 - p_1)^{\lfloor x \rfloor}$ for $x \geq 1$. Similarly,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 1, \\ 1 - (1 - p_2)^{\lfloor y \rfloor}, & \text{if } y \geq 1. \end{cases}$$

$\boxed{1, \text{B}}$

(ii) Then, for $z < 1$ we have $F_Z(z) = 0$ and for $z \geq 1$, we have $\boxed{\text{unseen} \Downarrow}$

$$\mathrm{P}(Z \leq z) = 1 - \mathrm{P}(Z > z) = 1 - \mathrm{P}(X > z, Y > z)$$
$$\overset{X,Y \text{ independent}}{=} 1 - \mathrm{P}(X > z)\mathrm{P}(Y > z)$$
$$= 1 - (1 - p_1)^{\lfloor z \rfloor}(1 - p_2)^{\lfloor z \rfloor} = 1 - \{1 - [1 - (1 - p_1)(1 - p_2)]\}^{\lfloor z \rfloor}.$$

Hence we deduced that Z follows the geometric distribution with parameter $1 - (1 - p_1)(1 - p_2)$. $\boxed{4, \text{C}}$

(d) We recall that the geometric distribution (as defined in the hint) arises when we count the number of independent Bernoulli trials needed to obtain the first success. Here we first need to wait until we see the first Heads appearing. This waiting time can hence be described by a geometric random variable X, say, with parameter 0.5 (since the coin is fair). After we have obtained Heads for the first time, we need to continue tossing the coin until we obtain Tails for the first time after having tossed Heads. Again, this waiting time can be described by a geometric random variable Y, say, with success probability 0.5. $\boxed{3, \mathrm{D}}$

We need to compute the mean of a geometric distribution:

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} x\mathrm{P}(X = x) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p = p\sum_{x=1}^{\infty} x(1-p)^{x-1}.$$

Note that $\boxed{1, \mathrm{B}}$

$$\sum_{x=1}^{\infty} x(1-p)^{x-1} = (-1)\frac{d}{dp}\sum_{x=0}^{\infty}(1-p)^{x}$$
$$\overset{\text{geom. series}}{=} (-1)\frac{d}{dp}\frac{1}{1-(1-p)} = -\frac{d}{dp}p^{-1} = p^{-2}.$$

Hence

$$\mathrm{E}(X) = p \times p^{-2} = p^{-1}.$$

Overall, we get that, on average, we need to wait for $\boxed{1, \mathrm{C}}$

$$E(X) + E(Y) = 2 + 2 = 4$$

tosses. $\boxed{1, \mathrm{A}}$

3. (a) We need that $f(x) \geq 0$ for all $x \in \mathbb{R}$ and that $\int_{-\infty}^{\infty} f(x)dx = 1$. seen $\Downarrow$

(b) (i) We have $\int_0^1 3xdx = \frac{3}{2}x^2\Big|_0^1 = \frac{3}{2}$. Hence, f is not a valid p.d.f.. However, if we choose $c = \frac{2}{3}$, then $cf(x) \geq 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} cf(x)dx = 1$. 2, A

meth seen $\Downarrow$

(ii) f is not a valid p.d.f. since it takes negative values. However, if we choose $c = -1$, then $cf(x) \geq 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} cf(x)dx = 1$. 2, A

(iii) f is not a valid p.d.f. since it takes negative values. Since it takes positive and negative values we cannot find a constant c which ensures that $cf(x) \geq 0$ for all $x \in \mathbb{R}$. 2, A

2, A

(c) (i) We compute meth seen $\Downarrow$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)dxdydz = \int_0^1\int_0^z\int_0^y cdxdydz = \int_0^1\int_0^z cydydz$$
$$= \int_0^1 \frac{c}{2}z^2dz = \frac{c}{6}.$$

Hence, we need to choose $c = 6$ to obtain a nonnegative function which integrates to 1. 3, D

unseen $\Downarrow$

(ii) We use the three-dimensional law of the unconscious statistician to conclude that

$$\mathrm{E}(XYZ) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyzf_{X,Y,Z}(x,y,z)dxdydz = 6\int_0^1\int_0^z\int_0^y xyzdxdydz$$
$$= 6\int_0^1\int_0^z \frac{1}{2}y^3zdydz = 3\int_0^1 \frac{1}{4}z^5dz = \frac{3}{4\times 6} = \frac{1}{8}.$$

3, D

seen $\Downarrow$

(d) (i) A partition of the sample space Ω is a collection $\{B_i : i \in \mathcal{I}\}$ (for a countable index set I) of disjoint events (meaning that $B_i \in \mathcal{F}$ and $B_i \cap B_j = \emptyset$ for $i \neq j$) such that $\Omega = \bigcup_{i\in\mathcal{I}} B_i$. 2, A
Example: $\Omega = \{0,1\}, \mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$ and partition $\{\{0\},\{1\}\}$. 1, A

seen $\Downarrow$

(ii) First we use the definition of the expectation, followed by the law of total probability:

$$\mathrm{E}(X) = \sum_x x\mathrm{P}(X=x) = \sum_x x\sum_{i\in\mathcal{I}}\mathrm{P}(X=x|B_i)\mathrm{P}(B_i)$$
$$= \sum_{i\in\mathcal{I}}\mathrm{P}(B_i)\sum_x x\mathrm{P}(X=x|B_i) = \sum_{i\in\mathcal{I}}\mathrm{P}(B_i)\mathrm{E}(X|B_i).$$

We use the fact that the series is absolutely convergent to justify that we are allowed to change the order of summation. 3, B

4. (a) (i) By a result in lectures (can be shown using moment generating functions) seen ⇓

$$\text{Each } X_i \text{ is normally distributed,}$$
$$\Rightarrow \text{Each } \frac{1}{n}X_i \text{ is normally distributed,}$$
$$\Rightarrow \overline{X} = \sum_{i=1}^{n}\left(\frac{1}{n}X_i\right) \text{ is normally distributed.}$$

Using the linearity of expectation, 1, A

$$\mathrm{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu.$$

Since the X_i are independent, using properties of the variance operator, 1, A

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

1, A

(It would also be acceptable to state that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$ by a result proved in lectures, since the $X_1, X_2, \ldots, X_n$ are independent and identically distributed.) Therefore, $\overline{X} \sim \mathbb{N}(\mu, \frac{\sigma^2}{n})$.

meth seen ⇓

(ii) Noticing that $Z = \frac{n-1}{n}S^2$ and using the linearity of expectation,

$$\mathrm{E}(Z) = \mathrm{E}\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\mathrm{E}\big(S^2\big) = \left(\frac{n-1}{n}\right)\sigma^2.$$

1, A

meth seen ⇓

(iii) Noticing that $Z = \frac{n-1}{n}S^2$ and using the property of the variance,

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2\mathrm{Var}\big(S^2\big) = \left(\frac{n-1}{n}\right)^2\frac{2\sigma^4}{n-1}$$

$$\Rightarrow \mathrm{Var}(Z) = \frac{2(n-1)\sigma^4}{n^2}.$$

1, A

sim. seen ⇓

(iv) The bias of an estimator $\widehat{\Theta}$ of a parameter θ is defined as $b_\theta(\widehat{\Theta}) = \mathrm{E}(\widehat{\Theta}) - \theta$. Therefore, using the result from (ii), the bias of Z estimating σ^2 is

$$b_{\sigma^2}(\widehat{Z}) = \mathrm{E}(Z) - \sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2 - \sigma^2 = \left(\frac{-1}{n}\right)\sigma^2.$$

1, A

(v) A theorem was proved in lectures which stated that, for any estimator $\widehat{\Theta}$ of a parameter θ, the mean squared error of $\widehat{\Theta}$ is

meth seen ⇓

$$\mathrm{E}[(\widehat{\Theta}-\theta)^2] = \left[b_\theta(\widehat{\Theta})\right]^2 + \mathrm{Var}(\widehat{\Theta}),$$

where $b_\theta(\widehat{\Theta})$ is the bias $\widehat{\Theta}$. Therefore, using the results from (iii) and (iv),

$$\begin{aligned}\mathrm{E}[(Z-\sigma^2)^2] &= [b_{\sigma^2}(Z)]^2 + \mathrm{Var}(Z)\\ &= \left[\left(\frac{-1}{n}\right)\sigma^2\right]^2 + \frac{2(n-1)\sigma^4}{n^2}\\ &= \frac{\sigma^4}{n^2} + \frac{(2n-2)\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}.\end{aligned}$$

2, A

sim. seen ⇓

(vi) Since the $X_1, X_2, \dots, X_n$ are i.i.d. $\mathbb{N}(\mu, \sigma^2)$, a theorem from lectures states that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$

where χ^2_{n-1} is a chi-squared distribution with $n-1$ degrees of freedom. Then, taking $b = \frac{n}{\sigma^2} > 0$,

$$bZ = \frac{nZ}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

2, B

sim. seen ⇓

(vii) Since the $X_1, X_2, \dots, X_n$ are i.i.d. $\mathbb{N}(\mu, \sigma^2)$, a theorem from lectures states that $\overline{X}$ and S^2 are independent. Therefore, using the (bi)linearity of the covariance,

$$\mathrm{Cov}(\overline{X}, Z) = \mathrm{Cov}(\overline{X}, \frac{n-1}{n}S^2) = \frac{n-1}{n}\mathrm{Cov}(\overline{X}, S^2) = \frac{n-1}{n} \times 0 = 0$$

since the covariance of two independent random variables is 0 (by another result in lectures).

2, B

(b) Fix a positive number $a > 0$, and define the random variable $\boxed{\text{seen} \Downarrow}$

$$Y_a = \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a. \end{cases}$$

This definition of Y_a ensures that $Y_a \leq X$ for all values of a and X. Therefore:

$$\mathrm{E}(Y_a) \leq \mathrm{E}(X).$$

$\boxed{2, \mathrm{B}}$

On the other hand, since Y_a is a discrete random variable, one can computes its expectation as

$$\mathrm{E}(Y_a) = 0 \times \mathrm{P}(X < a) + a \times \mathrm{P}(X \geq a).$$

$\boxed{1, \mathrm{A}}$

Combining the above equations, one obtains

$$a \times \mathrm{P}(X \geq a) \leq \mathrm{E}(X),$$

from which the Markov inequality follows. $\boxed{1, \mathrm{A}}$

$\boxed{\text{sim. seen} \Downarrow}$

(c) Although four of the five p-values in the table are below the threshold $\alpha = 0.01$, we need to account for the multiple testing and include a correction for multiple hypothesis testing. $\boxed{1, \mathrm{C}}$

Since there are 100 tests, if we use the Bonferroni correction the adjusted significance threshold would be $\alpha' = \alpha/100 = 0.0001 = 10^{-4}$. $\boxed{2, \mathrm{C}}$

Comparing the p-values in the table to this adjusted threshold, we see that only $2 \times 10^{-5} < \alpha'$ and $5 \times 10^{-6} < \alpha'$. Therefore, the research team should only conclude that two of the genetic variants, Genetic variant D and Genetic variant E, are significantly associated with the disease. $\boxed{1, \mathrm{C}}$

5. (a) $\boxed{\text{seen} \Downarrow}$

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - a)^2 &= \sum_{i=1}^{n}[(x_i - \overline{x}) + (\overline{x} - a)]^2 \\
&= \sum_{i=1}^{n}[(x_i - \overline{x})^2 + 2(x_i - \overline{x})(\overline{x} - a) + (\overline{x} - a)^2] \\
&= \sum_{i=1}^{n}(x_i - \overline{x})^2 + 2(\overline{x} - a)\sum_{i=1}^{n}(x_i - \overline{x}) + \sum_{i=1}^{n}(\overline{x} - a)^2 \\
&= \sum_{i=1}^{n}(x_i - \overline{x})^2 + 2(\overline{x} - a) \times 0 + n(\overline{x} - a)^2 \\
&= \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - a)^2,
\end{aligned}$$

where between the third line and the fourth line we used $\boxed{1, \text{A}}$

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

Since $n(\overline{x} - a)^2 \geq 0$, $\boxed{1, \text{A}}$

$$\sum_{i=1}^{n}(x_i - a)^2 \geq \sum_{i=1}^{n}(x_i - \overline{x})^2,$$

with equality only when $\overline{x} = a$, which proves the result. $\boxed{2, \text{A}}$

(b) Define $G(b_0, b_1) := \sum_{i=1}^{n}[y_i - b_0 - b_1 x_i]^2$. Then $\boxed{\text{seen} \Downarrow}$

$$G(b_0, b_1) = \sum_{i=1}^{n}[y_i - (b_0 + b_1 x_i)]^2 = \sum_{i=1}^{n}[(y_i - b_1 x_i) - b_0]^2.$$

$\boxed{1, \text{A}}$

Using 2 (a), this is minimised when

$$b_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - b_1 x_i) = \overline{y} - b_1 \overline{x}.$$

$\boxed{2, \text{A}}$

Returning to the expression for $G(b_0, b_1)$, and substituting in the value we have just found for b_0,

$$\begin{aligned}
G(b_0, b_1) &= \sum_{i=1}^{n}[(y_i - b_1 x_i) - b_0]^2 \\
&= \sum_{i=1}^{n}[(y_i - b_1 x_i) - (\overline{y} - b_1 \overline{x})]^2 \\
&= \sum_{i=1}^{n}[(y_i - \overline{y}) - b_1(x_i - \overline{x})]^2 \\
&= \sum_{i=1}^{n}[(y_i - \overline{y})^2 - 2b_1(x_i - \overline{x})(y_i - \overline{y}) + b_1^2(x_i - \overline{x})^2] \quad = S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx}.
\end{aligned}$$

Completing the square, $\boxed{1, \text{A}}$

$$\begin{aligned}
G(b_0, b_1) &= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} \\
&= S_{xx}\left(b_1^2 - 2b_1 \frac{S_{xy}}{S_{xx}} + \frac{S_{yy}}{S_{xx}}\right) \\
&= S_{xx}\left(b_1^2 - 2b_1 \frac{S_{xy}}{S_{xx}} + \left(\frac{S_{xy}}{S_{xx}}\right)^2 + \frac{S_{yy}}{S_{xx}} - \left(\frac{S_{xy}}{S_{xx}}\right)^2\right) \\
&= S_{xx}\left(\left[b_1 - \frac{S_{xy}}{S_{xx}}\right]^2 + \frac{S_{yy}}{S_{xx}} - \left(\frac{S_{xy}}{S_{xx}}\right)^2\right) \\
&= S_{xx}\left[b_1 - \frac{S_{xy}}{S_{xx}}\right]^2 + S_{yy} - \frac{(S_{xy})^2}{S_{xx}}
\end{aligned}$$

which shows that $G(b_0, b_1)$ is minimised when $b_1 = \frac{S_{xy}}{S_{xx}}$. Therefore, the parameter values that minimise $G(b_0, b_1)$ are $\boxed{2, \text{B}}$

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1 \overline{x} = \overline{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\overline{x}, \\
\widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}}.
\end{aligned}$$

$\boxed{1, \text{A}}$

Note that a calculus-based solution is also acceptable, provided it proves that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ minimise $G(b_0, b_1)$ globally.

(c) (i) The null hypothesis is that the means μ_X and μ_Y are equal, i.e. $\boxed{\text{seen} \Downarrow}$

$$H_0 : \mu_X = \mu_Y.$$

$\boxed{1, \text{A}}$

(ii) In order to use the t-test with full theoretical justification, it must be assumed that the random variables X and Y follow normal distributions and that $\text{Var}(X) = \text{Var}(Y)$. $\boxed{2, \text{B}}$

(The t-test can be applied in the case that $\text{Var}(X) \neq \text{Var}(Y)$ by using an approximation, but then the test is not exact.)

$\boxed{\text{unseen} \Downarrow}$

(d) The probability density function $f(x_i|\theta)$ for observation x_i, where $i \in \{1, 2, \dots n\}$, has the form

$$f(x_i|\theta) = \begin{cases} \frac{1}{\theta}, & \text{if } 0 \leq x_i \leq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Since the random variables $X_1, X_2, \dots, X_n$ are independent, the joint probability density function $f(\mathbf{x}|\theta)$ of $X_1, X_2, \dots, X_n$ can be written as

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } 0 \leq x_i \leq \theta, \text{ for all } i \in \{1, 2, \dots n\}, \\ 0, & \text{otherwise.} \end{cases}$$

$\boxed{2, \text{D}}$

Since the joint likelihood $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$, the maximum likelihood estimate of θ must be a value of θ such that (i) $x_i \leq \theta$ for $i \in \{1, 2, \dots n\}$ and (ii) this value maximises $1/\theta^n$ among all possible values for θ.

Since $1/\theta^n$ is a decreasing function of θ, the maximum likelihood estimate will be the smallest value of θ such that $x_i \leq \theta$ for $i \in \{1, 2, \dots n\}$.

Therefore, the maximum likelihood estimate of θ is $\widehat{\theta} = \max\{x_1, x_2, \dots, x_n\}$, and the maximum likelihood estimator is $\widehat{\theta} = \max\{X_1, X_2, \dots, X_n\}$. $\boxed{4, \text{D}}$

6. (a) Confidence intervals: meth seen $\Downarrow$

(i) Since $\overline{Y} \sim N(\mu, \frac{\sigma^2}{n})$, and for $Z \sim N(0,1)$ it is given in the table that $\mathrm{P}(-1.645 < Z < 1.645) \approx 0.90$, we can construct a 90% confidence interval for μ as

$$(\overline{y} - 1.645\frac{\sigma}{\sqrt{n}}, \overline{y} + 1.645\frac{\sigma}{\sqrt{n}}) = (5 - 1.645\frac{3}{\sqrt{10}}, 5 + 1.645\frac{3}{\sqrt{10}}).$$

2, A

(ii) If we cannot assume the variance is known, we can use the t-distribution with the sample variance, which in this case is given as $s^2 = 4$. 1, A

Using the table, we need to look up the value for $10-1 = 9$ degrees of freedom. The table gives this value as 2.262 (since for a 95% confidence interval, $\mathrm{P}(-2.262 < T < 2.262) \approx 0.95$). Therefore, a 95% confidence interval is

$$(\overline{y} - 2.262\frac{s}{\sqrt{n}}, \overline{y} + 2.262\frac{s}{\sqrt{n}}) = (5 - 2.262\frac{2}{\sqrt{10}}, 5 + 2.262\frac{2}{\sqrt{10}}).$$

1, B

(iii) In the case we do not know the distribution, we can use Chebyshev's inequality which states that for any $k > 0$,

$$\mathrm{P}(|\overline{Y} - \mu| < k\frac{\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow \left(\overline{Y} - k\frac{\sigma}{\sqrt{n}} < \mu < \overline{Y} + k\frac{\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}.$$

Taking $k = 10$, we have the 99% confidence interval

$$(5 - 10\frac{4}{\sqrt{10}} < \mu < 5 + 10\frac{4}{\sqrt{10}})$$

$$= (5 - 4\sqrt{10} < \mu < 5 + 4\sqrt{10}).$$

2, C

(b) (i) Each ϵ_i is assumed to follow a normal distribution with mean 0 and unknown variance σ^2, for $i \in \{1, 2, \ldots, n\}$, and the ϵ_i are also assumed to be independent. In other words, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ follows a joint (multivariate) Gaussian distribution with mean $\mathbf{0} = (0, 0, \ldots, 0)^T$ and covariance

seen ⇓

1, B

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix}$$

(i.e. a covariance matrix with each entry on the diagonal equal to σ^2 and all other entries 0).

sim. seen ⇓

(ii) Model 1 does **not** appear to fit the data well.

1, B

The reason is that there is clearly trend in the residuals (or: they appear to follow a "U"-shape), and they do not appear to be independently normally distributed.

1, C

Model 2 appears to fit the data well.

1, B

The reason is that the residuals appear to be normally distributed around 0, and they appear to be independent of each other.

1, B

(c) Since the random variables Z_i are independent and since $\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$, the joint probability density function of $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ given θ is $\boxed{\text{seen} \Downarrow}$

$$f(\mathbf{z}|\theta) = \prod_{i=1}^{n} f(z_i|\theta) = \prod_{i=1}^{n} \theta \exp(-\theta z_i) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} z_i\right) = \theta^n \exp(-\theta n \overline{z}),$$

for $z_1, z_2, \ldots, z_n > 0$, otherwise $f(\mathbf{z}|\theta) = 0$.

Note that this is the same as the likelihood of θ given the data $\mathbf{z}$. The posterior probability density function is proportional to the product of the prior and the likelihood: $\boxed{1, \text{A}}$

$$\pi(\theta|\mathbf{z}) \propto f(\mathbf{z}|\theta)\pi(\theta) = \theta^n \exp(-\theta n \overline{z}) \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)$$

$\boxed{1, \text{C}}$

$$\propto \theta^{n+\alpha-1} \exp(-\theta(n\overline{z} + \beta))$$

$$\propto \left(\frac{(n\overline{z} + \beta)^{n+\alpha}}{\Gamma(n+\alpha)}\right) \theta^{n+\alpha-1} \exp(-\theta(n\overline{z} + \beta)).$$

This last expression is the probability density function of a $\Gamma(n + \alpha, n\overline{z} + \beta)$ distribution, which shows that the posterior distribution of θ is a $\Gamma(n + \alpha, n\overline{z} + \beta)$ distribution. $\boxed{1, \text{C}}$

(d) It can be shown (for any two random variables X and Y) that $\boxed{\text{unseen} \Downarrow}$

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

This follows from the definition of covariance and was also shown in a problem sheet. Therefore, $\mathrm{Cov}(X,Y) = 4 - 2 \times 3 = -2$. $\boxed{2, \mathrm{D}}$

The correlation between X and Y, denoted ρ_{XY}, is defined as

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

It was also proved in lectures that $-1 \leq \rho_{XY} \leq 1$. Therefore,

$$\frac{[\mathrm{Cov}(X,Y)]^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = (\rho_{XY})^2 \leq 1.$$

$\boxed{1, \mathrm{D}}$

Therefore,

$$\mathrm{Var}(X) \geq \frac{[\mathrm{Cov}(X,Y)]^2}{\mathrm{Var}(Y)} = \frac{4}{\mathrm{Var}(Y)}.$$

Since Y is bounded on the interval $[2,5]$, it is a theorem in the notes (proved in Problem Sheet 8), that $\mathrm{Var}(Y) \leq \frac{(5-2)^2}{4} = \frac{9}{4}$, which implies $\boxed{1, \mathrm{D}}$

$$\begin{aligned} \frac{1}{\mathrm{Var}(Y)} &\geq \frac{4}{9} \\ \Rightarrow \mathrm{Var}(X) &\geq \frac{4}{\mathrm{Var}(Y)} \geq 4 \times \frac{4}{9} = \frac{16}{9}. \end{aligned}$$

Therefore, the standard deviation of X is bounded below by $\frac{4}{3}$, i.e. $\sqrt{\mathrm{Var}(X)} \geq \frac{4}{3}$.

$\boxed{2, \mathrm{D}}$

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