

This document contains seven test-style questions, each question being presented in Numbas-only submission style, and Canvas submission style. In *your* test, you will have a mix of questions from Chapters 3–6 of the course notes, some requiring all answers to be submitted directly within Numbas and some questions requiring more detailed calculations and an upload of calculations and/or code to Canvas. In *your* test, you will have questions totalling 50 marks. All questions will be accessed from within the Numbas application. Note that the example questions in this document are indicative of the style of questions you are likely to see, and to give you a feel for how you might gain marks. It is unlikely that questions in *your* test will be exactly the same as those presented here, and you should make sure you use all examples provided in the practical worksheets, and in the course notes, in your revision.

1 Test-style question on Transformations

A version of a question to be completed entirely within Numbas [5 marks]

Suppose $X \sim N(0, 1)$ and that $Y = X^2$. Which option, A–E, correctly specifies the probability density function (PDF) of Y ? Note that you are not required to use RStudio for this question.

- A. $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}$, for $0 \leq y < \infty$.
- B. $f_Y(y) = \frac{1}{2\pi} e^{-\frac{1}{4}y^2}$, for $0 \leq y < \infty$.
- C. $f_Y(y) = \frac{1}{\sqrt{8\pi y}} e^{-\frac{1}{2}y}$, for $0 \leq y < \infty$.
- D. $f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$, for $0 \leq y < \infty$.
- E. $f_Y(y) = \lambda e^{(y-\lambda e^y)}$, for $-\infty < y < \infty$.

Notes to students

- This question would appear within Numbas, to be completed entirely within Numbas by simply selecting the correct option (A–E)
- You would have to work out the PDF for Y by hand, on scrap paper, and then simply select the correct option within Numbas
- You would therefore not receive any marks for working, just for the correct answer
- There would be lots of different versions of the question in circulation (e.g. different distributions for X and different transformations), so it would be unlikely that two students get exactly the same question

A version of this question requiring solutions to be submitted to Canvas [10 marks]

Suppose $X \sim N(0, 1)$ and that $Y = X^2$. Find the probability density function (PDF) of Y . Include your full hand-written/typed solution in your solutions document that you will upload to Canvas. Note that you are not required to use RStudio for this question.

Notes to students

- This question would appear in Numbas, but would require you to upload your full solution (hand-written or typed) to Canvas
- The question will be marked by hand, and marks will be available for method *and* correct answer
- A photograph of your hand-written solution would be acceptable, provided it is clear and your working legible

2 Test-style question on Bivariate Distributions

A version of a question to be completed entirely within Numbas [15 marks]

Suppose $X \sim \text{Exp}(2)$ and that

$$f_{Y|X}(Y|X = x) = \begin{cases} \frac{e^{-y/x}}{x}, & \text{when } y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

In RStudio write a function to generate 500 pairs of realizations of (X, Y) .

- Produce a scatterplot of X against Y . Which plot, from the plots shown below, matches most closely to the plot you have produced in RStudio? (10 marks)
- What is your sample correlation between X and Y ? Enter your answer correct to 4 decimal places. (5 marks)

Note that you are not required to submit any R code code for this question.

Notes to students

- In this question, you would not upload your code anywhere
- You would use RStudio to write a function, produce a plot and compute a correlation coefficient
- For part (a), you would simply select the scatterplot that most closely resembles the plot you have produced; in part (b), you would enter your correlation coefficient
- There would be many versions of this question, and the parameters within (e.g. $\lambda = 2$ in this question) could also be randomised, so it is very unlikely that two students would get the same question

A version of this question requiring solutions to be submitted in Canvas [20 marks]

Suppose $X \sim \text{Exp}(2)$ and that

$$f_{Y|X}(Y|X = x) = \begin{cases} \frac{e^{-y/x}}{x}, & \text{when } y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

In RStudio write a function to generate 500 pairs of realizations of (X, Y) . Produce a scatterplot of X against Y , and obtain the sample correlation between X and Y . In your solutions document that you upload to Canvas, include (i) the R code for your main function (8 marks); (ii) your scatterplot (8 marks); (iii) the correlation coefficient (4 marks). No other code or solutions are needed.

Notes to students

- This question would appear in Numbas, but would require you to upload the code for your main function – along with the scatterplot and your correlation coefficient – to Canvas
- You should simply copy-and-paste the code from RStudio into MS WORD. Alternatively, a screenshot of your code in RStudio would suffice. No R files are needed
- The question would be marked by hand, and partial credit might be achieved for an incorrect function

3 Test-style question on Bivariate Distributions

A version of a question to be submitted to Canvas [25 marks]

Suppose

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 8 & 2 \\ 2 & 5 \end{pmatrix} \right].$$

The random variable Z is defined by $Z = X + 3Y$.

- (a) What is the distribution of Z ? State this distribution fully, including the name of the distribution and the value of any parameter(s) within the distribution.
- (b) Use R to simulate 1,000 pairs of observations on (X, Y) . Using this sample:
 - (i) Estimate $P(Z > 8)$. How does this compare to the true probability?
 - (ii) Estimate the correlation between X and Y . How does this compare to the true correlation?

In the solutions document that you upload to Canvas, you should include your working for part (a) (5 marks); the code you have used to produce your sample in part (b) (8 marks); your estimate of the probability in part (b)(i) (3 marks), with a comment on how it compares to the true value (3 marks); your estimate of the correlation in part (b)(ii) (3 marks), with a comment on how this sample value compares to the true value (3 marks) .

Notes to students

- This question would appear in Numbas, but would require you to include all necessary pencil-and-paper working, R code, and answers from RStudio, in the solutions document you upload to Canvas
- The question would be marked by hand, and partial credit might be achieved for incorrect R code or calculations
- Within Numbas the coefficients in Z , and the values in the variance-covariance matrix, could be randomised to ensure all students get different versions of the question

A version of this question to be submitted entirely within Numbas [15 marks]

Suppose

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 8 & 2 \\ 2 & 5 \end{pmatrix} \right].$$

The random variable Z is defined by $Z = X + 3Y$.

- (a) Use R to simulate 1,000 pairs of observations on (X, Y) . Which plot, from those shown below, most resembles a plot of your sample for X against that for Y ? (7 marks)
- (b) Use your sample in (a) to estimate $P(Z > 8)$, and enter this correct to 4 decimal places. (4 marks)
- (c) Use your sample in part (b) to estimate the correlation between X and Y , and enter this correct to 4 decimal places. (4 marks)

4 Test-style question on the Inverse CDF Method

A version of a question to be submitted to Canvas [20 marks]

It is suggested that the survival times of patients (X years) receiving a new cancer drug will follow a Weibull distribution with cumulative distribution function (CDF)

$$F_X(x; \lambda, \kappa) = 1 - e^{-(x/\lambda)^\kappa}, \quad x \geq 0, \quad \lambda, \kappa > 0,$$

where λ is referred to as the scale parameter and κ is the shape parameter.

- (a) Write an R function that uses the inverse CDF method to generate a sample of size N from the Weibull distribution. Your function should have *three* input variables: N , λ and κ . Include this function in your solutions document that you will upload to Canvas. (10 marks)
- (b) Past studies from a similar drug indicate that $\lambda = 2$ and $\kappa = 1.5$. Use these values, and your function in part (a), to simulate 10,000 realisations from X , and plot a histogram of these values. Colour the histogram orange, and make sure your axes are appropriately labelled. Include this histogram in your solutions document for upload to Canvas. (6 marks)
- (c) Use your simulations in part (b) to estimate the probability that a patient receiving this drug will survive for at least 5 years. (4 marks)

A version of this question to be submitted entirely within Numbas [10 marks]

It is suggested that the survival times of patients (X years) receiving a new cancer drug will follow a Weibull distribution with cumulative distribution function (CDF)

$$F_X(x; \lambda, \kappa) = 1 - e^{-(x/\lambda)^\kappa}, \quad x \geq 0, \quad \lambda, \kappa > 0,$$

where λ is referred to as the scale parameter and κ is the shape parameter.

- (a) Write an R function that uses the inverse CDF method to generate a sample of size N from the Weibull distribution. Your function should have *three* input variables: N , λ and κ . There is nothing to submit to Numbas for this part of the question.
- (b) Past studies from a similar drug indicate that $\lambda = 2$ and $\kappa = 1.5$. Use these values, and your function in part (a), to simulate 10,000 realisations from X . Use your simulated values to estimate the probability that a patient receiving this drug will survive for at least 5 years. Enter your answer correct to 4 decimal places. (10 marks)

5 Test-style question on Transformations

A version of a question to be submitted to Canvas [10 marks]

If $X \sim \text{Gamma}(\alpha, \lambda)$ then it has probability density function

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha, \lambda > 0,$$

where $\Gamma(A) = (A-1)!$ for integer-valued A . Also, $E(X) = \alpha/\lambda$ and $\text{Var}(X) = \alpha/\lambda^2$.

- (a) Suppose $X \sim \text{Gamma}(4, 1)$. Find the probability density function (PDF) for $Y = X/5$.
Make sure you include all necessary working in your solutions document. (7 marks)
- (b) Find $E[Y]$. (3 marks)

A version of this question to be submitted entirely within Numbas [12 marks]

If $X \sim \text{Gamma}(\alpha, \lambda)$ then it has probability density function

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha, \lambda > 0,$$

where $\Gamma(A) = (A-1)!$ for integer-valued A . Also, $E(X) = \alpha/\lambda$ and $\text{Var}(X) = \alpha/\lambda^2$.

- (a) Suppose $X \sim \text{Gamma}(4, 1)$. Find the probability density function (PDF) for $Y = X/5$, and evaluate this to four decimal places at the points in the table below (9 marks):

Y	1	1.5	2.5
$f_Y(y)$			

- (b) Find $E[Y]$, entering your answer correct to four decimal places. (3 marks)

6 Test–style question on the Inverse CDF method

A version of a question to be submitted to Canvas [20 marks]

A glaciologist is studying the annual rate of retreat, Y metres, of the *Zachariae Isstrom* glacier in northern Greenland. Experience suggests that the Pareto distribution, with probability density function

$$f_Y(y; \alpha, \lambda) = \frac{\alpha \lambda^\alpha}{y^{\alpha+1}}, \quad \alpha, \lambda > 0, \quad y > \lambda,$$

might be a reasonable model to use for the rate of retreat of the glacier, where α and λ are shape and rate parameters, respectively.

- (a) Data from the *Zachariae Isstrom* glacier are only available for the years 1991–2022. Write an R function that simulates a series of N annual glacial retreat values for this glacier using the inverse CDF method. Include this function in your solutions document. Also include your workings.
- (b) We know from studies of a nearby glacier that $\alpha \approx 1.5$; the minimum annual glacial retreat we might expect at the *Zachariae Isstrom* glacier is 5.5 metres. Using your function in part (a), produce a boxplot of a sample of $N = 10^6$ values. Make sure you remove the outliers from the plot.
- (c) Using your sample in part (b), find the probability that in 2023 the *Zachariae Isstrom* glacier will retreat by more than 12 metres.

A version of this question to be submitted entirely within Numbas [12 marks]

A glaciologist is studying the annual rate of retreat, Y metres, of the *Zachariae Isstrom* glacier in northern Greenland. Experience suggests that the Pareto distribution, with probability density function

$$f_Y(y; \alpha, \lambda) = \frac{\alpha \lambda^\alpha}{y^{\alpha+1}}, \quad \alpha, \lambda > 0, \quad y > \lambda,$$

might be a reasonable model to use for the rate of retreat of the glacier, where α and λ are shape and rate parameters, respectively.

- (a) Write an R function that uses the inverse CDF method to generate a sample of size N from the Pareto distribution. Your function should have *three* input variables: N , α and λ . There is nothing to submit to Numbas for this part of the question.
- (b) Past studies from a similar drug indicate that $\alpha = 1.5$ and $\lambda = 5.5$. Use these values, and your function in part (a), to simulate 10^6 realisations from Y . Use your simulated values to estimate the probability that in 2023 the *Zachariae Isstrom* glacier will retreat by more than 12 metres (give your answer to four decimal places). (10 marks)

7 Test–style question on simulating random variables

A version of a question to be submitted to Canvas [15 marks]

Suppose we suspect a coin is biased. We assume

$$P(\text{Head}) = p \sim \text{Beta}(2, 5),$$

where R's built-in command `rbeta(N, a, b)` can be used to generate a sample of size N from the Beta distribution with parameters a and b .

- (a) Use R to generate a sample of 10,000 values for p . In your solutions file that you will upload to Canvas, include:
 - (i) The code you have used to produce this sample;
 - (ii) A histogram of your sample, coloured yellow;
 - (iii) Your sample mean, and sample variance, of p .
 - (iv) Is our belief that the coin is biased in favour of heads or tails? Explain your answer.
- (b) Now suppose you wish to perform an experiment in which the coin is flipped 10 times and the number of heads counted. Rather than perform this experiment by hand, you will use R to simulate the experiment.
 - (i) Suppose X : Number of heads observed, and that $X|p \sim \text{Bin}(10, p)$. Write a function in R to simulate 10^6 observations on X . Include this function in your solutions document that you will upload to Canvas.
 - (ii) Using your function in (i), produce a bar plot for X . Include this in your solutions document.
 - (iii) Using your sample, find $P(X > 5)$, and comment.

A version of this question to be submitted entirely within Numbas [10 marks]

Suppose we suspect a coin is biased. We assume

$$P(\text{Head}) = p \sim \text{Beta}(2, 5),$$

where R's built-in command `rbeta(N, a, b)` can be used to generate a sample of size N from the Beta distribution with parameters a and b .

- (a) Use R to generate a sample of 10,000 values for p . From this sample,
 - (i) estimate $E(p)$;
 - (ii) estimate $\text{Var}(p)$;
 - (iii) decide whether you think the coin is biased in favour of heads or tails (choose one).

- (b) Now suppose you wish to perform an experiment in which the coin is flipped 10 times and the number of heads counted. Rather than perform this experiment by hand, you will use R to simulate the experiment. Suppose X : Number of heads observed, and that $X|p \sim \text{Bin}(10, p)$. Write a function in R to simulate 10^6 observations on X .
- (i) Produce a barplot of your sample. Which of the barplots shown below most closely resembles your histogram?
 - (ii) Use your sample to find $P(X > 5)$, to four decimal places.