Durham University

EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2026	MATH2687-WE01

Title:
Data Science and Statistical Computing II

Time:	2 hours	
Additional Material provided:	Statistical tables	
Materials Permitted:	None	
Calculators Permitted:	Yes	Models Permitted: Casio FX83 series or FX85 series.

Instructions to Candidates:	Answer all questions. The indicative marks shown in brackets for the main parts of each question are given as a guide to the weighting the markers expect to apply. Write your answer in the white-covered answer booklet with barcodes. Begin your answer to each question on a new page.

Revision:	

ED01/2026

SECTION A

1. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ consist of a set of n independent observations of a discrete random variable, X, having unknown probability distribution. We are interested in estimating some real-valued parameter θ using a statistic $S(\cdot)$.

(a) Write down the estimator $\hat{\theta}$ and give detailed steps for how to estimate its variance via the Bootstrap. **[4]**

(b) The parameter of interest, θ, is the probability that X is zero. Write down the equation for a statistic, $S(\cdot)$, to estimate this parameter.

You collect data:

$$x_1 = 0, x_2 = 4, x_3 = 1.$$

Assuming you perform Bootstrap resampling, write down all possible values for $S(\cdot)$ and the probability of observing each one. **[3]**

(c) Find $\mathbb{E}[\bar{S}^\star]$ without performing any simulation and estimate the implied bias in using this statistic to estimate θ. **[3]**

---

2. (a) Let the random variable X have probability density function (pdf):

$$f(x) = \begin{cases} (x+1)^{-2} & \text{if } x \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}.$$

Simulate three values from this pdf via inverse transform sampling using the following values simulated from the Uniform$(0, 1)$ distribution:

$$0.56, \quad 0.85, \quad 0.26$$ **[4]**

(b) Let the random variable X have probability mass function (pmf) parameterised by $K \in \mathbb{N}$:

$$\mathbb{P}(X = x) = f(x) = \begin{cases} \frac{1}{K} & \text{if } x = 1 \\ \frac{1}{x(x-1)} & \text{if } x \in \{2, \dots, K\} \\ 0 & \text{otherwise} \end{cases}.$$

Let $K = 4$.

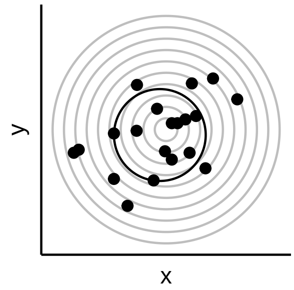
(i) Sketch the cumulative distribution function (cdf) **[3]**

(ii) Simulate three values from this pmf via inverse transform sampling using the following values simulated from the Uniform$(0, 1)$ distribution:

$$0.49, \quad 0.91, \quad 0.75$$ **[3]**

SECTION B

3. In the science of ballistics, the Circular Error Probability (CEP) is used to evaluate the precision (i.e., spread, not accuracy) of firearms. It is defined to be the radius of a circle, centred on the mean of all shots, which is expected to include 50% of bullets fired. For example:



By making simplifying assumptions, we can directly model the random radial distance of each shot, R, from the mean as Rayleigh distributed with probability density function:

$$f(r) = \begin{cases} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) & \text{when } r \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}.$$

for some $\sigma > 0$.

The Department of Defence needs to test if the manufacturer's claim of a CEP of 0.5 is plausible given 20 shots at a test target. Those 20 shots resulted in an empirical CEP radius of 0.64.

They ask you to perform the test:

$$H_0 : \text{CEP} = 0.5 \text{ versus } H_1 : \text{CEP} > 0.5$$

(a) Find the cdf of the Rayleigh distribution. Show that the median of the Rayleigh distribution is $\sigma\sqrt{2\log 2}$. Find the value of σ which corresponds to the null hypothesis to be tested. **[4]**

(b) In order to perform the above test, your colleague suggests using the test statistic

$$T = h(R_1, \ldots, R_{20}) = \text{median}(R_1, \ldots, R_{20}) - 0.5.$$

Would this be ok and why (or why not)? **[1]**

(c) Describe in detail how to conduct a Monte Carlo hypothesis test in this particular setting (eg state what is simulated etc). **[5]**

Question continues on the next page

(d) Following the steps you provided, your colleague has produced 100 simulations of the test statistic, ordered and shown as follows:

$$
\begin{array}{ccccc}
-0.159, & -0.145, & -0.136, & -0.133, & -0.127, \\
-0.120, & -0.115, & -0.115, & -0.111, & -0.110, \\
 & & \ldots \text{ 80 other values } \ldots & & \\
0.115, & 0.117, & 0.120, & 0.129, & 0.130, \\
0.134, & 0.152, & 0.164, & 0.168, & 0.179.
\end{array}
$$

Estimate the p-value based on this (small) Monte Carlo simulation. What would you tell the Department of Defence? **[2]**

(e) Define the resampling risk (do not try to actually calculate its value). **[3]**

Exam continues on the next page

4. We saw in tutorials that it is easy to inverse transform sample random variables, X, having probability density function (pdf):

$$f(x \mid \alpha) = \begin{cases} \alpha x^{\alpha-1} & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

for any $\alpha > 0$.

(a) Your friend didn't do the tutorial and so immediately tried rejection sampling instead. They considered using a Uniform(0,1) distribution as proposal. For what values of α will this work? **[2]**

(b) Consider only the set of values for which the Uniform(0,1) distribution is a valid proposal. As a function of α, what is the probability that any single iteration of the rejection sampler will be accepted? **[2]**

For the rest of this question we are interested in the behaviour of an estimator $\hat{\mu}_n$ of the parameter $\mu := \mathbb{E}\left[-\sqrt{X}\right]$ when $\alpha = 2$.

(c) Assume you have n Monte Carlo simulations $\{x_1, \ldots, x_n\}$ from the above pdf with $\alpha = 2$. Write down the equation for the Monte Carlo estimator, $\hat{\mu}_n$, and calculate the variance of $\hat{\mu}_n$ as a function of n. **[3]**

(d) Your friend suggests that it would be more efficient to use an importance sampler to compute $\hat{\mu}_n$, using Monte Carlo simulations from a proposal pdf with the same form, but with a different parameter value α, say $\alpha = a \neq 2$. Calculate (in terms of a) the variance of such an importance sampling estimator for $\hat{\mu}_n$, where the target still has $\alpha = 2$, and state the valid range of values for a. **[5]**

(e) For what range of values of a does the importance sampling estimator have lower variance than the standard Monte Carlo estimator? **[3]**
