Total: 20 points, Time: 40 minutes

Provide justification for all solutions unless otherwise stated.

Question 1

(i) **(2 points)** State Markov's inequality.

(ii) **(2 points)** Using Markov's inequality, prove Chebyshev's inequality:

If X is a random variable with finite mean μ and finite variance σ^2, then for all $c > 0$,

$$\mathrm{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}.$$

(iii) **(3 points)** Suppose Y is a random variable following an unknown distribution with unknown mean μ and unknown variance σ^2, although we will assume that μ and σ^2 are both finite. Showing all working, compute a lower bound for the probability that Y is within 3 standard deviations of its mean.

Question 2

(2 points) Suppose you are given a data set containing the ages (in years, i.e. positive integers) of the approximately $n = 1000$ students in the Department of Mathematics at Imperial. State which plot you would use to visualise the distribution of the ages, and provide motivation for your choice.

Question 3

(5 points) Suppose $X_1, X_2, \ldots, X_n$, where $n = 100$, are independent and identically distributed random variables following a normal distribution with mean μ and variance σ^2. The value of μ is unknown, but σ^2 is known to be $\sigma^2 = 10$. We will also assume that the value of μ is finite. Suppose we observe $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and that we compute the sample mean of these observations to be $\overline{x} = 4.2$. Using Table 1 below, construct a 95% confidence interval for μ, providing full justification.

Question 4

Suppose that the $n > 2$ random variables $X_1, X_2, \ldots, X_n$ are independent and each follows the same distribution which has finite mean μ and finite variance σ^2. Furthermore, assume it is known that the fourth (raw) moment is $\mathrm{E}[X_i^4] = \mu^4 + 4\mu^2\sigma^2 + 2\sigma^4$ for $i = 1, 2, \ldots, n$. We decide to define $\widehat{\Theta}$, an estimator of the variance σ^2, as:

$$\widehat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Clearly stating any results or properties used:

(i) **(2 points)** Compute $b_{\sigma^2}(\widehat{\Theta})$, the bias of $\widehat{\Theta}$.

(ii) **(4 points)** Compute the mean squared error of $\widehat{\Theta}$.

Table 1: 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
2.576	0.995