MATH7501 Exercise sheet 8 — to be done by Thursday 15th March

THIS SET OF EXERCISES IS THE FOURTH ASSESSMENT

To assist the marking of this assessment, please write your name CLEARLY at the top of each sheet of paper, and indicate clearly (e.g. by underlining) the final answer to each part of a question. You need to submit your MATH7501 coursework in person by 4:00pm on Thursday at Room 610 in the Department of Mathematics. Please ensure that your ID card is checked when you submit your coursework.

All solutions must be your own work.

1. $X_1, \ldots, X_n$ are independent observations from a normal distribution with mean μ and variance σ^2. It can be shown that the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 ,$$

where $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$, is distributed as $\Gamma(\alpha, \lambda)$ with $\alpha = (n-1)/2$ and $\lambda = (n-1)/2\sigma^2$.

   (a) Using the results for the gamma distribution in your notes, show that s^2 is an unbiased estimator of σ^2. Also, state its variance.

   (b) Consider a statistic of the form $T_k = ks^2$, for some constant k. Find expressions for the bias and variance of T_k as an estimator of σ^2. Hence, or otherwise, find the value of k that minimises the mean squared error of T_k as an estimator of σ^2.

6 marks

2. (a) Z is a standard normal random variable, and U is distributed as χ^2_ν. Z and U are independent. State the distribution of $T = Z/\sqrt{U/\nu}$. What is the distribution of T^2?

   (b) Y is a random variable distributed as $F_{1,5}$. Use your answer to part (a), together with tables of the t (NB **NOT** the F!) distribution, to find $P(Y > 4)$. Also, find the value of c such that $P(Y > c) = 0.01$.

4 marks

3. A DIY store sells timber strips, each with a nominal length of 2 metres. A sample of 16 strips is measured, and the recorded lengths (in cm) are as follows:

201.8	197.8	202.0	198.0	197.9	200.2	199.6	200.2
201.4	203.3	199.2	198.2	198.3	199.7	200.5	197.9

   (a) Prepare a stem-and-leaf diagram representing these data. Also, calculate their sample mean and variance.

   (b) From long experience, it is known that the true standard deviation of timber strips is 1cm. Assuming that strip lengths have a normal distribution, calculate a 95% confidence interval for the mean of this distribution. Do you think the mean is equal to its nominal value of 2 metres? Justify your answer.

   (c) Is the assumption of normality reasonable here? State any other assumptions that underly your answer to part (b).

10 marks