Research School of Finance, Actuarial Studies and Statistics

EXAMINATION

Semester 1 - End of Semester, 2016

STAT2001/6039 Introductory Mathematical Statistics/Principles of Mathematical Statistics

Examination/Writing Time Duration: 180 minutes
Reading Time: 15 minutes

Exam Conditions:
Central Examination
Students must return the examination paper at the end of the examination
This examination paper is not available to the ANU Library archives

Materials Permitted In The Exam Venue:
(No electronic aids are permitted e.g. laptops, phones)
No restriction

Materials to Be Supplied To Students:
1 x 20 page plain
Scribble Paper

Instructions to Students:

1. This exam paper has 4 pages. Please ensure your paper has the correct number of pages.
2. The exam includes a total of 6 problems. The problems are of unequal value, with marks indicated for each problem. STAT6039 students should attempt to solve all 6 problems. STAT2001 students should attempt all problems except Problem 6.
3. The maximum number of marks attainable in the exam is 160 for STAT2001 students and 180 for STAT6039 students.
4. This exam will count toward 60% or 80% of your final assessment for the course.
   It will count 80% if and only if you do better in it than in the mid-semester exam.
5. Do not round calculations until providing your final answer to each problem. Work using five significant digits. Final numerical solutions should be rounded to three significant digits (e.g. 0.00217) or expressed as simple fractions (e.g. 7/12).
6. Include all workings for each problem, as marks will not be awarded for solutions that do not include workings.
7. Ensure that you include your student number on your answer book.

Problem 1 ***[20 marks in total]***

(a) Suppose that $X \sim Beta(2,1)$.
Find and sketch the density of $Y = X / (X+1)$. Also calculate EY. ***[10 marks]***

(b) Suppose that $U \sim U(0,1)$, $Z \sim N(0,1)$ and $U \perp Z$.
Find and sketch the density of $R = Z / U$. Also calculate $P(R > 8)$. ***[10 marks]***

Problem 2 ***[40 marks in total]***

A fair die is rolled repeatedly. Define a "swipe" as the occurrence of 1, 2 and 3 in a row, in any order. For example, the sequence 1532531234 contains two swipes, namely 312 and 123. The first of these swipes occurs on the 8th roll, and the second occurs on the 9th roll.

(a) Find the expected number of the roll on which the first swipe occurs. ***[10 marks]***

(b) Find the probability that the first swipe occurs on an even-numbered roll. ***[10 marks]***

(c) Find the minimum number of rolls that result in at least one expected swipe. ***[10 marks]***

(d) Find the probability that six rolls result in no swipes. ***[10 marks]***

Problem 3 ***[30 marks in total]***

The country of *Tigeria* has many millions of people.

The proportion of females in the country is q, and all the other people are males.

There is interest in p, the proportion of people in Tigeria who have the disease *criplea*.

Every person in Tigeria has the same probability of having criplea (namely p), independently of everyone else and regardless of gender.

A sample is obtained from the population of Tigeria by sampling people randomly, independently, and one by one, until w women have been sampled.

(a) Find formulae for the mean and variance of the number of persons in the sample with criplea. Then evaluate these formulae if w = 200, q = 0.6 and p = 0.05. ***[10 marks]***

(b) Suppose that w = 1, q = 0.35, p = 0.02 and nobody in the sample has criplea. Find the density, distribution and mean of the number of people in the sample. ***[10 marks]***

(c) Suppose that 70% of Tigeria's population is female, the sample has five persons, and two of these five have criplea. Find the maximum likelihood estimates of w and p. ***[10 marks]***

Problem 4 ***[40 marks in total]***

Consider a random sample of size n from a normal distribution with an unknown positive mean μ and with variance $\sigma^2 = \mu^2$.

We are interested in μ and are considering three very different estimators of this quantity, defined as follows:

1. $\bar{y}$, the sample mean
2. $\hat{\mu}$, the maximum likelihood estimator of μ
3. $\hat{\sigma} = cs$, where c is some constant and s is the sample standard deviation.

(a) Find a formula for the value of c which ensures that $\hat{\sigma}$ is an unbiased estimator of μ. Then calculate this value of c for the case where n = 6. ***[10 marks]***

(b) Find a formula for the efficiency of $\bar{y}$ relative to $\hat{\sigma}$, where $\hat{\sigma}$ is unbiased for μ. Then calculate this efficiency for the case where n = 6. ***[10 marks]***

(c) Find a formula for $\hat{\mu}$. Then calculate $\hat{\mu}$ for when $y = (y_1,...,y_n)$ = (1.2,1.7, 0.1). Then illustrate this value of $\hat{\mu}$ in a sketch of the likelihood function. ***[10 marks]***

(d) We wish to test the null hypothesis that $\mu = 3$ against the alternative hypothesis that $\mu > 3$ using s^2 as the test statistic, a rejection region of the form (k,∞) and a significance level of α = 0.05. If the sample size is n = 2, find the value of k and derive a formula for the power function. Evaluate this function at μ = 2, 3, 4 and 10. Then sketch this function and mark in its values at μ = 2, 3, 4 and 10. ***[10 marks]***

Problem 5 ***[30 marks in total]***

Consider a discrete random variable with density

$$f(y) = \frac{1}{3} \times 2^{-|y|},\ y = 0, \pm 1, \pm 2, \pm 3, ...$$

(a) Define $X = (Y \mid Y > 1)$ and then do the following.

- **(i)** Find the density of *X*, denoted $f_X(x) = P(X = x)$. ***[2 marks]***
- **(ii)** Find the mode of *X*, denoted *M* = *Mode*(*X*). ***[1 mark]***
- **(iii)** Find the median of *X*, denoted *m* = *Median*(*X*). ***[1 mark]***
- **(iv)** Find the mean of *X*, denoted $\mu = EX$. ***[2 marks]***
- **(v)** Find the variance of *X*, denoted $\sigma^2 = VX$. ***[2 marks]***
- **(vi)** Sketch $f_X(x)$ and mark in *M*, *m* and μ . ***[2 marks]***

(b) Repeat (a) but with $X = Y \times I(Y > 1)$. ***[10 marks]***

Note: Here, *I* denotes the standard indicator function, such that, for any event *E*, the function returns $I(E) = 1$ if *E* occurs, and it returns $I(E) = 0$ if *E* does not occur.

(c) Repeat (a) but with $X = |Y - 1|$. ***[10 marks]***

Problem 6 ***[to be done only by STAT6039 students]*** ***[20 marks in total]***

(a) Apply a simple linear regression to the data $(y_1, y_2, y_3) = (1,3,4)$ with covariates $(x_1, x_2, x_3) = (0,1,1)$. Estimate the intercept and slope parameters α and β . Predict *m*, an independent future *y* value with covariate 0.5. Construct a 95% prediction interval for *m*, assuming the error terms are independent and standard normal. ***[10 marks]***

(b) Suppose that $y_i = \alpha u_i + \beta x_i + e_i$, where the x_i and u_i values are known and the e_i values are independent and standard normal. Apply the method of least squares to derive estimates of α and β . Then calculate these estimates if $(u_1, u_2, u_3) = (1, 0, 2)$ and the corresponding x_i and y_i values are as given in (a). ***[10 marks]***

END OF EXAMINATION