Australian National University

RESEARCH SCHOOL OF FINANCE, ACTUARIAL STUDIES AND APPLIED STATISTICS

Final Examination – First Semester 2015

INTRODUCTORY MATHEMATICAL STATISTICS (STAT2001)

PRINCIPLES OF MATHEMATICAL STATISTICS (STAT6039)

Study Period: 15 minutes

Time Allowed: 3 hours

Permitted Material: No restrictions

- ***STAT2001*** *students should attempt the* ***first seven*** *problems (only). For these students, the exam is out of* ***115*** *marks.*
- ***STAT6039*** *Students should attempt* ***all eight*** *problems. For these students, the exam is out of* ***125*** *marks.*
- *Draw a box around each solution and express each numerical solution in decimal form to at least four significant digits (e.g.* $\boxed{0.007204}$*). Start your solution to each problem on a new page.*
- *To ensure full marks, show all the steps in working out your solutions. Marks may be deducted for not showing appropriate calculations or formulae, or for not clearly referencing the results in the text book or course material which you are using.*

Problem 1 ***[10 marks in total]***

Ann and Bob are about to play a game, as follows. They will take turns rolling a standard six-sided die until a 6 comes up or until 5 comes up three times in a row. The winner will be the last person to roll the die. Ann will begin the game by rolling first.

(a) Find the probability that Ann will win the game. ***[5 marks]***

(b) Find the expected total number of rolls of the die. ***[5 marks]***

Problem 2 ***[25 marks in total]***

Suppose that the numbers 3.1, 2.8, 2.4, 3.0, 3.7 constitute a random sample of size n = 5 from a distribution defined by the following probability density function:

$$f(y) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}(y-\lambda)}, \quad y \geq \lambda \quad (\lambda > 0).$$

(a) Find the method of moments estimate (MME) of λ. ***[5 marks]***

(b) Find the maximum likelihood estimate (MLE) of λ. ***[5 marks]***

(c) Find the mean squared error (MSE) of the sample mean,

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

(considered as an estimator of λ), and evaluate this MSE when λ = 0.6. ***[5 marks]***

(d) Find an unbiased estimate of λ by suitably modifying the MLE of λ in (b).

[5 marks]

(e) Calculate the efficiency of the estimator in (d) relative to the estimator in (a).

[5 marks]

Problem 3 ***[20 marks in total]***

Consider a single random variable

$$Y \sim Beta(k,1),$$

where k is an unknown positive constant.

(a) Find the method of moments estimate of k and the maximum likelihood estimate of k, assuming that the observed value of Y is $y = 0.7$. ***[5 marks]***

(b) Derive the density of $X = 1/Y^k$. Then evaluate this density at $x = 5$ if $k = 3.8$.

[5 marks]

(c) Using (b), or otherwise, find a central 95% confidence interval for k if $y = 0.7$.

[10 marks]

Problem 4 ***[20 marks in total]***

Consider two random variables

$$X, Y \sim iid\ U(0,1).$$

Then define:

$$T = 2X/(X+1)$$

$$U = (2Y-1)/X.$$

(a) Find and sketch the density of T. ***[5 marks]***

(b) Using (a), or otherwise, calculate ET. ***[5 marks]***

(c) Find and sketch the density of U. ***[10 marks]***

Problem 5 ***[10 marks in total]***

Consider a *large* random sample $Y_1,...,Y_n$ from a distribution with mean μ and variance σ^2, where both of these parameters are positive and finite. Assume that the underlying distribution is positive, in the sense that $P(Y_i > 0) = 1$ for all $i = 1,\ldots,n$.

We observe the sample mean $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$, and we know the sample size n.

Derive an approximate central $1-\alpha$ confidence interval (CI) for μ under each of the following assumptions:

(a) $\sigma^2 = \mu^2$ ***[5 marks]***

(b) $\sigma^2 = \mu$. ***[5 marks]***

In each case, calculate the approximate 95% CI for μ when $\bar{y}$ = 5 and n = 64.

Problem 6 ***[10 marks]***

Consider a distribution defined by a probability density function which has the form

$$f(y) = \frac{c}{(y+1)^2},\ 0 \le y \le \theta \qquad (\theta > 0) \qquad \text{(where } c \text{ is a constant).}$$

A single value y is observed from this distribution.

Derive an explicit formula for a suitable p-value for testing the null hypothesis that

$$\theta = k$$

(where k is a specified positive constant) against the alternative hypothesis that

$$\theta > k.$$

Then calculate q, the p-value if y = 4 and k = 5. Also find z, the value of y which would result in a p-value of 0.1 if k = 5. Then sketch the p-value as a function of y if k = 5. In your sketch, show the two points $(y, p) = (4, q)$ and $(y, p) = (z, 0.1)$.

Problem 7 ***[20 marks in total]***

A committee of 6 persons is selected from 9 men and 13 women – one person at a time, randomly and without replacement. (At each selection, all the persons who are not yet on the committee have the same chance of being chosen for the committee.)

(a) Find the (single) probability that the oldest man is on the committee *or* the oldest woman is *not* on the committee. ***[5 marks]***

(b) If at least one man is on the committee, find the probability that the first person chosen to be on the committee is a man. ***[5 marks]***

(c) A *second* committee, having 4 persons, is selected from the 9 men and 13 women – one person at a time, randomly, without replacement, and independently of the first committee (of 6). Find the mean and variance of Y, the number of men on *both* committees. (Hint: Let X be the number of men on the *first* committee.) ***[10 marks]***

Problem 8 **(To be attempted only by STAT6039 students)** ***[10 marks]***

Consider n = 1000 data pairs (x_i, y_i) for which the following are summary statistics:

$$\sum_{i=1}^{n} x_i = 2532.1, \qquad \sum_{i=1}^{n} x_i^2 = 8599.29$$

$$\sum_{i=1}^{n} y_i = -10.9, \qquad \sum_{i=1}^{n} y_i^2 = 1776.01, \qquad \sum_{i=1}^{n} x_i y_i = 1720.24.$$

Fit a simple linear regression to the dataset. Calculate point estimates of all the unknown parameters in the model. Calculate a 95% confidence interval (CI) for the slope parameter. Also predict v, a future and independent y-value with x-value u = 0.6. Finally, calculate a 95% prediction interval for v and also a 95% CI for the mean of v.

END OF EXAMINATION