OLLSCOIL NA hEIREANN, CORCAIGH
The National University of Ireland, Cork

COLAISTE NA hOLLSCOILE, CORCAIGH
University College, Cork

2015/2016

Semester 2 - Summer 2016

Second University Examination in Science
Financial Mathematics and Actuarial Science; Mathematical Sciences;
Higher Diploma in Statistics

ST2054 - Probability and Mathematical Statistics

Mr. P. Cooper
Professor F. O'Sullivan
Dr. M. Cronin
Mr. A. O'Leary

Time allowed: Three hours.

Statistical tables are available. A calculator may be used provided that it does not contain any information stored by any person prior to this examination.

Fifteen minutes of reading time are permitted prior to this examination.

PLEASE ANSWER ANY NINE QUESTIONS
All questions carry equal marks

PLEASE DO NOT TURN THIS PAGE UNTIL INSTRUCTED TO DO SO

PLEASE
ENSURE THAT YOU HAVE THE CORRECT EXAM PAPER

Question 1

(i) State the Axioms of Probability. [3 marks]

(ii) Using the Axioms of Probability, prove that for any two events A and B, the following expression is true:

$$P(A \cup B) \leq P(A) + P(B)$$

[2 marks]

(iii) Using the Axioms of Probability, prove by induction that the expression in **(ii)** can be extended as follows:

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

[2 marks]

(iv) Consider two events A and B with $P(A) = 0.4$ and $P(B) = 0.7$. Determine the maximum and minimum possible values of $P(A \cap B)$ and the conditions under which each of these values is attained. [3 marks]

Question 2

You are about to buy a car. Let G be the event that the car is in good condition and H be the event that the car salesman is honest. It is known that $P(G) = 0.7$. It is also known that the car will be in good condition with an honest salesman 90% of the time and with a dishonest salesman 15% of the time.

(i) Find the probability of getting an honest salesman. [4 marks]

(ii) A friend has suggested that $P(H|G) = P(G|H)$. Is this always true? If not, state when it is true. [3 marks]

(iii) You bought a car which turned out to be in good condition. What is the probability you bought it from an honest salesman? [3 marks]

Question 3

The random variable Y has the following probability density function (pdf):

$$f(y) = \begin{cases} \frac{(60y-6y^2)}{1000} & 0 \leq y \leq 10 \\ 0 & otherwise \end{cases}$$

(i) Find an expression for $P(Y < y | \, Y < 4)$. [4 marks]

(ii) Hence, or otherwise, show that the conditional pdf: [1 mark]

$$f_{Y|Y<4}(y|y<4) = \frac{15}{88}y - \frac{3}{176}y^2 \;\; for \;\; 0 \leq y < 4.$$

(iii) Calculate the conditional expected value $E[Y | \, Y < 4]$ and the conditional variance $Var[Y | \, Y < 4]$. [2 marks]

(iv) Find an expression for $P(Y < u | \, Y > 1 \, and \, Y < 4)$ [2 marks]

Question 4

Suppose that the probability that a family has n children is given by ap^n, $n \geq 1$, where $a \leq (1-p)/p$.

(i) Find the probability that a family has no children. [2 marks]

(ii) If each child is equally likely to be a boy or a girl (independently of other births), find an expression for the probability that a family consists of k boys (and any number of girls). [3 marks]

Consider a sequence of Bernoulli trials, where the probability of success (S) on each trial is p and the probability of failure (F) is $1-p$. Let X denote the completed number of successive trials with the same outcome before a different outcome occurs – the completed run length. For example $X = 3$ when you get either $SSSF$ or $FFFS$.

(iii) Find the probability density function (pdf) of X. [3 marks]

(iv) Hence, or otherwise, find $E[X]$. [2 marks]

Question 5

(i) Explain what is meant by *convergence in probability* for a sequence of random variables. [3 marks]

A carpenter must cut 100 lengths of wood of equal length, given by a metal bar which acts as a standard. The carpenter uses the metal standard to cut the first length of wood, and thereafter uses the most recently cut length of wood as his standard for the next cut. Each time he cuts a length of wood there is an error in the length compared with the standard he is using for that cut, and this error has mean $0mm$, and standard deviation $2mm$, and is independent of previous cuts.

Let Y_n be the random variable corresponding to the error between the n^{th} length of wood and the original standard metal bar.

Let X_i be the random variable corresponding to the error between the i^{th} and $(i-1)^{th}$ lengths of wood, with X_0 being the original standard metal bar.

(ii) Compute the expected value and variance of Y_{100}. [2 marks]

(iii) Citing the appropriate theorem, establish the distributions of $\bar{X}_{100}$, the sample mean of the error between cuts, and of $\bar{Y}_{100}$, the mean error after 100 cuts. [2 marks]

(iv) Find the probability that the difference in length between the metal standard and the final length cut will exceed $6mm$. [3 marks]

Question 6

The continuous random variables X and Y have a joint probability density function (pdf) $f(x,y)$ where:

$$f(x,y) = \begin{cases} f(x,y) = 2e^{-x-y} & 0 < x < y < \infty \\ 0 & otherwise \end{cases}$$

(i) Show that the marginal pdf of X is:

$$f_X(x) = 2e^{-2x}, \;\; x > 0$$

and the marginal pdf of Y is [3 mark]

$$f_Y(y) = 2e^{-y}(1 - e^{-y}), \;\; y > 0$$

(ii) Show that X and Y are dependent random variables. [3 marks]

(iii) Show that the conditional pdf of Y is [2 marks]

$$e^{x-y}, \;\; y > x$$

(iv) Find $P[Y < 1|X = 0.5]$. [2 mark]

Question 7

(i) Show that for any random variables X and Y, [3 marks]

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X]).$$

(ii) The number N of claims occuring on a collection of insurance policies has a Poisson distribution with parameter m. Let $Y_i \overset{i.i.d.}{\sim} \Gamma(\alpha, \beta)$ be the claim amount of the i^{th} claim. The random variables $\{Y_i\}$ are independent and identically distributed as $\Gamma(\alpha, \beta)$. Derive the moment generating function of X_N, the total amount of the aggregate claim. [4 marks]

(iii) Calculate $Var(X_N)$.(*Hint: use (i).*) [3 marks]

Question 8

Consider a random sample of n independent realizations of the random variable with distribution $N(\mu, \sigma^2)$.

(i) Show that $s^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$ is an unbiased estimator for σ^2. [3 marks]

Consider the following estimator for σ^2:

$$\tilde{s}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n+1}$$

(ii) Find the bias for $\tilde{s}^2$ as an estimator of σ^2. [2 marks]

(iii) Find the mean square error for $\tilde{s}^2$ as an estimator of σ^2. [5 marks]

Question 9

A random sample of size n is drawn from a Normal distribution with mean µ and variance σ^2. The following null and alternative hypotheses are being considered:

$$H_0 : \mu \leq 350 \;\; ; \;\; H_A : \mu > 350$$

(i) Explain what is meant by Type 1 and Type 2 errors. [2 marks]

(ii) Calculate the required sample size n so that the risk of making a Type 1 error is at most 2%, and also so that if the value of µ is at least 400, there is a 90% chance that the Null hypothesis would be rejected. If the population standard deviation is 100, find the required value of n. [5 marks]

(iii) A sample size of $n = 100$ was chosen. It was found that $\bar{X} = 370$ and $s = 80$. Calculate a 99% confidence interval for µ. [3 marks]

Question 10

Consider the one-way analysis of variance model with equal numbers of observations, $j = 1, 2, \ldots, n$, per treatment group $i = 1, 2, \ldots, t$:

$$Y_{ij} = \mu + \tau_i + e_{ij},, \quad i = 1, 2, \ldots, t \;\; j = 1, 2, \ldots, n$$

(i) Give an explanation of each of the terms in the model.

State the usual model assumptions. [2 marks]

(ii) Derive the least squares estimators for μ and τ_i . [3 marks]

An insurance company wishes to compare four methods (A, B, C and D) for training new employees. As part of this investigation, a test is given to 20 of these employees, of whom 5 have taken training method A, 5 training method B, 5 training method C and 5 training method D. The test scores are as follows.

Method	Test Score (y_{ij})					$\sum_j y_{ij}$	$\sum_j y_{ij}^2$
A	86	79	81	70	84	**400**	**32154**
B	90	76	88	82	89	**425**	**36265**
C	82	68	73	71	81	**375**	**28729**
D	73	80	82	74	78	**387**	**30013**

$$\sum_i \sum_j y_{ij} = 1587 \qquad \sum_i \sum_j y_{ij}^2 = 126711$$

(iii) Carry out a test of the hypothesis that the mean effectiveness of each of the 4 methods is the same. Clearly state your null and alternative hypotheses, along with your conclusions. [5 marks]

Formulae for ST2054

Law of Total Probability: $P(A) = \sum_{i=1}^{k} P(A|E_i)P(E_i)$ when $\cup_{i=1}^{k} E_i = \Omega$ and all E_i are mutually exclusive.

$$\begin{array}{lll} X \sim Binomial(n,p) & E(X) = np & Var(X) = np(1-p) \\ U \sim Chi\text{-}Square(n) & E(X) = n & Var(X) = 2n \\ X \sim N(\mu, \sigma^2) & E(X) = \mu & Var(X) = \sigma^2 \\ X \sim Poisson(m) & E(X) = m & Var(X) = m \end{array}$$

Probability Distributions and MGFs:

Exponential pdf $f(x) = \lambda e^{-\lambda x}$ **Uniform pdf** $f(x) = \frac{1}{b-a}$

Binomial pmf $P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$

Multinomial pmf $P[X_1 = x_1, \ldots X_k = x_k] = \frac{n!}{x_1!, \ldots, x_k!} p_1^{x_1} \ldots p_k^{x_k}$, $\sum_{i=1}^{k} x_i = n$

Poisson pmf $P[X = k] = \frac{m^k e^{-m}}{k!}$ $MGF = e^{m(e^s - 1)}$

Gamma pdf $f(x) = \frac{1}{\Gamma(\alpha)} (\frac{x}{\beta})^{\alpha-1} e^{\frac{-x}{\beta}} \frac{1}{\beta}$ $MGF = \frac{1}{(1-\beta s)^\alpha}$

Normal pdf $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $MGF = e^{s\mu + \frac{1}{2}s^2\sigma^2}$

Chebyshev's Inequality:

$p[|X - \mu| > k\sigma] \leq \frac{1}{k^2}$, where μ and σ^2 are the mean and variance of X.

Linear Regression

$\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Analysis of Variance:

$SST = (\sum_i \sum_j (y_{ij} - \bar{y})^2) = \sum_i \sum_j (y_{ij})^2 - \frac{(y_{..})^2}{n}$ $SSB = [\sum_i \frac{(y_{i.})^2}{n_i}] - \frac{(y_{..})^2}{n}$

Where $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$ and $y_{..} = \sum_i \sum_j y_{ij}$