BIRKBECK

(University of London)

BSc EXAMINATION FOR INTERNAL STUDENTS
SCHOOL OF BUSINESS, ECONOMICS AND INFORMATICS

Probability and Statistics
EMMS098S5

Thursday, 30th May, 2013
Afternoon, 2.30 pm – 5.30 pm

This examination contains two sections: Section A and Section B. Questions in Section A are worth 5 *marks each and questions in Section B are worth* 20 *marks each.*
Candidates should attempt **all** *of the questions in Section A and* **two** *questions from the four in Section B.*
New Cambridge Statistical Tables are provided.
Candidates can use their own calculator, provided the model is on the circulated list of authorized calculators or has been approved by the chair of the Mathematics & Statistics Examination Sub-board.

Please turn over

©Birkbeck College, 2013

Section A

1. Let A and B be two events such that $\Pr(A) = 0.5$, $\Pr(A \cup B) = 0.9$ and $\Pr(B) = p$.

   (a) What is the value of p if A and B are mutually exclusive? [2]

   (b) What is the value of p if A and B are statistically independent? [3]

2. A company receives 40% of its orders over the internet. On a given day, 18 independently placed orders are received.

   (a) State the probability distribution of the number of orders received over the internet on the given day, including the values of any parameters. [2]

   (b) Evaluate the expected number of the orders received over the internet. [1]

   (c) Evaluate, correct to 3 decimal places, the standard deviation of the number of orders received over the internet. [1]

   (d) Calculate, correct to 3 decimal places, the probability that the number of the orders received over the internet is 10 or more. [1]

Please turn over

©Birkbeck College, 2013

3. According to research carried out by an insurance company, 0.13% of the population is involved in a certain type of accident each year. Its 8000 policy holders may be regarded as randomly selected from the population.

    (a) State the probability distribution that should be used for predictions of the number of policy holders who will have an accident of the given type next year, including the value of any parameter. [2]

    (b) Write down, correct to 3 decimal places, the probability that no more than 5 of the policy holders have an accident of this type next year. [1]

    (c) Calculate, correct to 3 decimal places, the probability that more than 10 of the policy holders have an accident of this type next year. [1]

    (d) Calculate, correct to 3 decimal places, the probability that between 6 and 10, inclusive, of the policy holders have an accident of this type next year. [1]

4. A company manufactures concrete blocks whose weights are normally distributed with mean 12.0 kg and standard deviation 0.2 kg.

    (a) Find the probability, correct to 4 decimal places, that the weight of a randomly chosen block exceeds 12.1 kg. [2]

    (b) If 10 such blocks are selected at random, find the probability, correct to 4 decimal places, that the total weight of the blocks exceeds 121 kg. [3]

Please turn over

©Birkbeck College, 2013

5. A study was conducted to assess the amounts of chemical residues found in the brain tissue of pelicans of various age-groups, including juveniles and nestlings. The chemical residues are assumed to accumulate over time. Random samples of 10 juveniles and 13 nestlings were taken and the quantity of DDT in parts per million was measured for each sample member.

   For the juveniles, the sample mean was 0.041 and the sample standard deviation 0.019. For the nestlings, the sample mean was 0.026 and the sample standard deviation 0.016.

   It is assumed that the data from the juveniles may be regarded as a random sample from a normal distribution with mean μ_1 and variance σ^2 and the data from the nestlings as a random sample from a normal distribution with mean μ_2 and variance σ^2.

   (a) Calculate the pooled estimate of the variance σ^2. [1]

   (b) Calculate the appropriate test statistic to test the null hypothesis that $\mu_1 = \mu_2$ against the alternative hypothesis that $\mu_1 > \mu_2$, and draw conclusions. [4]

6. In a random sample of 200 voters in a particular state, 87 say that they are definitely going to vote for Candidate A as president.

   (a) Find a 95% confidence interval for the proportion of all voters in the state who would say that they are definitely going to vote for Candidate A. [2]

   (b) Calculate the size of the random sample that would need to be taken to ensure that a 95% confidence interval was at most 0.02 in length, whatever the sample proportion. Give your answer correct to the nearest hundred. [3]

Please turn over

7. The birthdays of a random sample of 200 students in a college were found to fall in the quarters of the year as follows:

Quarter	1	2	3	4
Frequency	62	48	44	46

(a) Write down the expected frequencies under the null hypothesis they that the students are drawn from a population in which birthdays are uniformly distributed throughout the year. [1]

(b) Calculate a corresponding test statistic and state its approximate distribution under the null hypothesis. [2]

(c) Calculate the p-value of the test statistic and draw conclusions. [2]

8. A random sample of 200 voters with low incomes and a random sample of 100 voters with high incomes has been taken to determine their attitude towards a certain piece of legislation that has been proposed. The data are given in the table below.

	low income	high income
for the legislation	116	37
against the legislation	84	63

Carry out an analysis to investigate whether there is significant evidence of any association between income category and attitude towards the legislation. State your conclusions. [5]

Please turn over

©Birkbeck College, 2013

Section B

9. (a) i) Given two events, A and B, in some sample space S, define the conditional probability, $\Pr(A|B)$, stating under what condition it is defined.
Assuming that the conditional probabilities are defined, prove that

$$\Pr(B|A) = \frac{\Pr(B)\Pr(A|B)}{\Pr(A)}.$$

[4]

ii) Given some sample space S, let A be any event and let $B_1, B_2, \ldots, B_k$ be a collection of pairwise mutually exclusive and exhaustive events, none of which has probability zero. Assuming that $\Pr(A) > 0$, state Bayes' Theorem. [2]

(b) A medical investigator is studying three drugs identical in appearance, numbered 1, 2 and 3, respectively. When the drugs are injected into guinea pigs, the probability that an antitoxin will develop is 1/4 for Drug 1, 1/8 for Drug 2 and 1/3 for Drug 3. There are two bottles of Drug 1, three bottles of Drug 2 and one bottle of Drug 3. The investigator, in hurrying to give a demonstration of these drugs, takes one bottle at random without noting its number. In transit the label is torn off. When a guinea pig is injected with this drug, **no** antitoxin forms.

Evaluate to three decimal places the probability that it was Drug 2 that was injected. [7]

(c) A masked attacker has, after a violent struggle, murdered a young man. Police arrest a man who is on their records because of a minor offence that he committed in the past and whose DNA matches that of blood obtained from the victim's clothing. It is assessed that the probability that the DNA of a randomly chosen man would match the DNA found on the victim is 1 in 20 million. It is further assessed that there is a pool of 200,000 men, any of whom, in the absence of any other evidence, might have carried out the murder.

Adopting a Bayesian approach, and explaining your reasoning, evaluate to three decimal places the probability that the man arrested is guilty of the murder. [7]

Please turn over

10. Consider a continuous random variable X with probability density function f given by

$$f(x) = x \exp\left(-\frac{x^2}{2}\right) \qquad (x \geq 0).$$

(a) Using the formula for the standard normal density function, show that

$$\int_0^\infty \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{\frac{\pi}{2}}.$$

[2]

(b) Using integration by parts, or otherwise, find an expression for the mean of the random variable X. [4]

(c) Let μ_k denote the kth moment of X, i.e., $\mu_k = E(X^k)$ $(k \geq 0)$. Using integration by parts, show that

$$\mu_k = k\mu_{k-2} \qquad (k \geq 2).$$

[4]

(d) Deduce that

$$\mathrm{var}(X) = \frac{4-\pi}{2}\ .$$

[3]

(e) Find an expression for the cumulative distribution function $F(x)$ of X. [2]

(f) Find an expression for the median of the distribution of X.

Comment on the relative values of the median and the mean with reference to the shape of $f(x)$. [5]

Please turn over

©Birkbeck College, 2013

11. In an experiment to investigate whether there was significant evidence of an underlying difference between the mean measurements given by two types of caliper, Caliper 1 and Caliper 2, the diameter of a ball bearing was measured by 12 inspectors, each using the two types of caliper. The results are given in the table below.

Diameter of ball bearing in cm.

Inspector	Caliper 1	Caliper 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

In the Minitab output on the next page, two alternative test procedures are carried out, a parametric one and a nonparametric one.

(a) In the case of the parametric test:

i) State precisely the statistical model that is being used, defining carefully any notation that you use. Specify the null and alternative hypotheses in terms of the model parameters. [5]

ii) Write down a general formula for the test statistic that is used in the parametric test and state its distribution under the null hypothesis. [2]

iii) Draw conclusions in the present case. [2]

(b) In the case of the nonparametric test:

i) State any distributional assumptions that are made and specify the null and alternative hypotheses. [2]

ii) Using the tabulated data, show exactly how the Wilcoxon statistic has been calculated. [6]

(c) Comment on the essential difference between the assumptions that underlie the parametric and the nonparametric tests, and on how the conclusions that may be drawn from the two procedures compare in the present case. [3]

. . . continued

Please turn over

```
MTB > Paired 'Caliper1' 'Caliper2'.

Paired T-Test and CI: Caliper1, Caliper2

Paired T for Caliper1 - Caliper2

             N      Mean     StDev   SE Mean
Caliper1    12  0.266250  0.001215  0.000351
Caliper2    12  0.266000  0.001758  0.000508
Difference  12  0.000250  0.002006  0.000579


95% CI for mean difference: (-0.001024, 0.001524)
T-Test of mean difference = 0 (vs not = 0): T-Value = 0.43  P-Value = 0.674

MTB > Let c4 = 'Caliper1'-'Caliper2'
MTB > Name c4 'Diff'
MTB > WTest 0.0 'Diff';
SUBC>   Alternative 0.

Wilcoxon Signed Rank Test: Diff

Test of median = 0.000000 versus median not = 0.000000

          N for   Wilcoxon         Estimated
       N   Test  Statistic      P     Median
Diff  12      8       21.5  0.674  0.0005000
```

Please turn over

©Birkbeck College, 2013

12. A haemacytometer is used to count the numbers of yeast cells present in the different squares when a microscope slide is partitioned into 400 equally sized squares. If the yeast cells are distributed randomly over the plate, the numbers of yeast cells per square should follow a Poisson distribution. The observed frequency distribution of the number of yeast cells in each square is tabulated below.

No. of yeast cells in square	0	1	2	3	4	5	6	7	8	9	total
Observed frequency	75	103	121	54	30	13	2	1	0	1	400

(a) i) In the following Minitab output a goodness-of-fit test has been carried out. State the null hypothesis that is being tested, write down a general formula for the test statistic that is being used, and state its approximate distribution under the null hypothesis. [3]

ii) Draw conclusions in the present case. [2]

```
MTB > PGoodness 'Cells';
SUBC>   Frequencies 'Freq';
SUBC>   RTable.
Goodness-of-Fit Test for Poisson Distribution
Data column: Cells
Frequency column: Freq
Poisson mean for Cells = 1.8

                       Poisson             Contribution
Cells  Observed    Probability  Expected     to Chi-Sq
0            75       0.165299    66.120       1.19272
1           103       0.297538   119.015       2.15507
2           121       0.267784   107.114       1.80024
3            54       0.160671    64.268       1.64056
4            30       0.072302    28.921       0.04028
5            13       0.026029    10.411       0.64358
>=6           4       0.010378     4.151       0.00551

  N  N*  DF   Chi-Sq  P-Value
400   0   5  7.47796    0.187
```

(b) i) Explain why the amalgamation of frequencies into a category "≥ 6" has taken place. [2]

ii) Carry out a further amalgamation of the last two categories in the Minitab output into a category "≥ 5", calculating the corresponding observed frequency, Poisson probability, expected frequency and contribution to the chi-square statistic. Calculate the corresponding value of the test-statistic for testing the null hypothesis, state the degrees of freedom, and use the tables to evaluate the corresponding p-value to 3 decimal places. [8]

. . . continued

Please turn over

(c) In the following output, as an alternative to the goodness-of-fit test, the dispersion test has been carried out.

  i) State the property of the moments of the Poisson distribution that underlies the dispersion test. [1]

  ii) Write down a general formula for the index of dispersion that is used as the test statistic. [1]

  iii) In the present case, state the value of the index of dispersion, its approximate distribution under the null hypothesis, and its p-value, and draw conclusions. [3]

```
MTB > Let k1 = sum(Cells*Freq)/400
MTB > Name k1 'mean'
MTB > Print 'mean'.
Data Display
mean    1.80000

MTB > Let k2 = sum(Cells*Cells*Freq) - 400*mean*mean
MTB > Name k2 'SS'
MTB > Let k3 = SS/399
MTB > Name k3 'variance'
MTB > Print 'variance'
Data Display
variance    1.96491

MTB > Let k4 = SS/mean
MTB > Name k4 'index'
MTB > CDF 'index' k5;
SUBC> ChiSquare 399.
MTB > Let k6 = 1 - k5
MTB > Name k6 'p-value'
MTB > Print  'index' 'p-value'
Data Display
index      435.556
p-value    0.100282
```