Faculty of Actuaries Institute of Actuaries

EXAMINATION

7 September 2006 (am)

Subject CT3 — Probability and Mathematical Statistics Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*
2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*
3. *Mark allocations are shown in brackets.*
4. *Attempt all 12 questions, beginning your answer to each question on a separate sheet.*
5. *Candidates should show calculations where this is appropriate.*

Graph paper is required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator.

1 A bag contains 8 black and 6 white balls. Two balls are drawn out at random, one after the other and without replacement.

Calculate the probabilities that:

(i) The second ball drawn out is black. [1]

(ii) The first ball drawn out was white, given that the second ball drawn out is black. [2]

[Total 3]

2 Let A and B denote independent events.

Show that A and $\overline{B}$, the complement of event B, are also independent events. [3]

3 Consider 12 independent insurance policies, numbered 1, 2, 3, …, 12, for each of which a maximum of 1 claim can occur. For each policy, the probability of a claim occurring is 0.1.

Find the probability that no claims arise on the group of policies numbered 1, 2, 3, 4, 5 and 6, and exactly 1 claim arises in total on the group of policies numbered 7, 8, 9, 10, 11, and 12. [3]

4 In a large portfolio 65% of the policies have been in force for more than five years. An investigation considers a random sample of 500 policies from the portfolio.

Calculate an approximate value for the probability that fewer than 300 of the policies in the sample have been in force for more than five years. [3]

5 In a random sample of 200 policies from a company's private motor business, there are 68 female policyholders and 132 male policyholders.

Let the proportion of policyholders who are female in the corresponding population of all policyholders be denoted π.

Test the hypotheses

$$H_0: \pi = 0.4 \ \text{ v } \ H_1: \pi < 0.4$$

stating clearly the approximate probability value of the observed statistic and your conclusion. [4]

6 It is assumed that claims arising on an industrial policy can be modelled as a Poisson process at a rate of 0.5 per year.

(i) Determine the probability that no claims arise in a single year. [1]

(ii) Determine the probability that, in three consecutive years, there is one or more claims in one of the years and no claims in each of the other two years. [2]

(iii) Suppose a claim has just occurred. Determine the probability that more than two years will elapse before the next claim occurs. [2]

[Total 5]

7 A commuter catches a bus each morning for 100 days. The buses arrive at the stop according to a Poisson process, at an average rate of one per 15 minutes, so if X_i is the waiting time on day i, then X_i has an exponential distribution with parameter $\frac{1}{15}$ so

$$E[X_i] = 15, Var[X_i] = 15^2 = 225.$$

(i) Calculate (approximately) the probability that the total time the commuter spends waiting for buses over the 100 days exceeds 27 hours. [3]

(ii) At the end of the 100 days the bus frequency is increased, so that buses arrive at one per 10 minutes on average (still behaving as a Poisson process). The commuter then catches a bus each day for a further 99 days. Calculate (approximately) the probability that the total time spent waiting over the whole 199 days exceeds 40 hours. [3]

[Total 6]

8 Let $\bar{X}_1$ denote the mean of a random sample of size n from a normal population with mean μ and variance σ_1^2, and let $\bar{X}_2$ denote the mean of a random sample also of size n from a normal population with the same mean μ but with variance σ_2^2. The two samples are independent.

Define W as the weighted average of the sample means

$$W = \alpha\bar{X}_1 + (1-\alpha)\bar{X}_2$$

(i) Show that W is an unbiased estimator of μ. [1]

(ii) Obtain an expression for the mean square error of W. [2]

(iii) Show that the value of α for which W has minimum mean square error is given by

$$\alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

and verify that the optimum corresponds to a minimum. [3]

(iv) Consider the special case when the variances of the two random samples are equal to a common value σ^2. State (*do not derive*) the maximum likelihood estimator of μ calculated from the combined samples, and compare it with the estimator obtained in (iii). [2]

[Total 8]

9 (i) Show that for continuous random variables X and Y:

$$E[Y] = E[E(Y|X)].$$ [3]

(ii) Suppose that a random variable X has a standard normal distribution, and the conditional distribution of a Poisson random variable Y, given the value of $X = x$, has expectation $g(x) = x^2 + 1$.

Determine $E[Y]$ and $Var[Y]$. [5]

[Total 8]

10 Let $(X_1, X_2, \ldots, X_n)$ be a random sample of a gamma(4.5, λ) random variable, with sample mean $\bar{X}$.

(i) (a) Using moment generating functions, show that $2\lambda n\bar{X} \sim \chi^2_{9n}$.

(b) Construct a 95% confidence interval for λ, based on $\bar{X}$ and the result in (i)(a) above.

(c) Evaluate the interval in (i)(b) above in the case for which a random sample of 10 observations gave a value $\sum x_i = 21.47$. [9]

(ii) (a) Show that the maximum likelihood estimator of λ is given by

$$\hat{\lambda} = \frac{4.5}{\bar{X}}.$$

(b) Show that the asymptotic standard error of $\hat{\lambda}$ is estimated by

$$\frac{\hat{\lambda}}{(4.5n)^{1/2}}.$$

(c) Construct a 95% confidence interval for λ based on the asymptotic distribution of $\hat{\lambda}$, and evaluate this interval in the case for which a random sample of 100 observations gave a value $\sum x_i = 225.3$.

[9]

[Total 18]

11 A survey of financial institutions which offered tax-efficient savings accounts was conducted. These accounts had limits on the amounts of money that could be invested each year. The study was interested in comparing the maturity values achieved by investing the maximum possible amount each year over a certain time period.

The following values (in units of £1,000 and rounded to 2 decimal places) are the maturity values for such investments for 8 high street banks and 12 other banks (i.e. those without high street branches).

High street banks (x_1): 11.91, 11.87, 11.83, 11.66, 11.53, 11.49, 11.49, 11.42

$$(\Sigma x_1 = 93.20,\ \Sigma x_1^2 = 1{,}086.0470)$$

Other banks (x_2): 12.23, 12.17, 12.16, 11.90, 11.82, 11.77, 11.74, 11.70, 11.64, 11.60, 11.55, 11.50

$$(\Sigma x_2 = 141.78,\ \Sigma x_2^2 = 1{,}675.8224)$$

(i) Draw a diagram in which the maturity values for high street and other banks may be compared. [2]

(ii) Calculate a 95% confidence interval for the difference between the means of the maturity values for high street and other banks, and comment on any implications suggested by the interval. [6]

(iii) (a) Show that a test of the equality of variances of maturity values for high street banks and other banks is not significant at the 5% level.

(b) Comment briefly on the validity of the assumptions required for the interval in (ii). [4]

The following values (in units of £1,000 and rounded to 2 decimal places) are the maturity values for the maximum possible investment for a random sample of 12 building societies (a different kind of financial institution).

Building societies (x_3): 12.40, 12.19, 12.06, 12.01, 12.00, 11.97, 11.94, 11.92, 11.88, 11.86, 11.81, 11.79

$$(\Sigma x_3 = 143.83,\ \Sigma x_3^2 = 1{,}724.2449)$$

(iv) Add further points to your diagram in part (i) such that the maturity values for all three types of financial institution may be compared. [1]

(v) Use one-way analysis of variance to compare the maturity values for the three different types of financial institution, and comment briefly on the validity of the assumptions required for analysis of variance. [6]

(vi) Interpret the results of the statistical analyses conducted in (ii) and (v). [2]

[Total 21]

12 A development engineer examined the relationship between the speed a vehicle is travelling (in miles per hour, mph), and the stopping distance (in metres, m) for a new braking system fitted to the vehicle. The following data were obtained in a series of independent tests conducted on a particular type of vehicle under identical road conditions.

Speed of vehicle (x):	10	20	30	40	50	60	70
Stopping distance (y):	5	10	23	34	40	54	75

$$\sum x = 280 \quad \sum y = 241 \quad \sum x^2 = 14{,}000 \quad \sum y^2 = 11{,}951 \quad \sum xy = 12{,}790$$

(i) Construct a scatterplot of the data, and comment on whether a linear regression is appropriate to model the relationship between the stopping distance and speed. [4]

(ii) Calculate the equation of the least-squares fitted regression line. [5]

(iii) Calculate a 95% confidence interval for the slope of the underlying regression line, and use this confidence interval to test the hypothesis that the slope of the underlying regression line is equal to 1. [5]

(iv) Use the fitted line obtained in part (ii) to calculate estimates of the stopping distance for a vehicle travelling at 50 mph and for a vehicle travelling at 100 mph.

Comment briefly on the reliability of these estimates. [4]

[Total 18]

END OF PAPER