Imperial College
London

Department of Mathematics

BSc and MSci EXAMINATIONS (MATHEMATICS) MAY–JUNE 2005

This paper is also taken for the relevant examination for the Associateship.

M1S Probability and Statistics I

DATE: Friday, 13th May 2005 TIME: 10 am – 12 noon

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used. Statistical tables will not be available.

1.

(a) State the three axioms of probability for events defined on a sample space Ω.

(b) **(i)** Using the axioms of probability prove that for any event $E \subseteq \Omega$,

$$\mathbf{P}(E') = 1 - \mathbf{P}(E).$$

(ii) Hence determine which of the following events are most likely to occur:

(a) at least one six in a simultaneous throw of four dice.

(b) at least one double six in two throws of a pair of dice.

(c) **(i)** State, without proof, De Morgan's first law for events $F_1, \ldots, F_n$.

(ii) Hence show that, if $F_1, \ldots, F_n$ are independent, then

$$\mathbf{P}\left(\bigcup_{i=1}^{n} F_i\right) = 1 - \prod_{i=1}^{n}[1 - \mathbf{P}(F_i)].$$

2.

(a) **(i)** What do we mean by saying that events $F_1, \ldots, F_n$ form a partition of the sample space Ω?

Suppose $\mathbf{P}(F_i) > 0$ for $i = 1, \ldots, n$. Let E be any event in Ω, with $\mathbf{P}(E) > 0$.

(ii) State the Theorem of Total Probability.

(iii) Derive Bayes' formula for $\mathbf{P}(F_i|E)$.

(b) Footballers are routinely tested for the use of performance-enhancing drugs. A player provides two blood samples, the first of which is then tested. If this test is positive, indicating that performance-enhancing drugs are present, the second sample is tested. If both blood tests are positive, then the player has failed the overall test.

Suppose that a player is selected at random, and two blood samples (regarded as identical) are obtained. Let events T_1 and T_2 correspond respectively to the events that the first and second samples test positive, and let D be the event that performance-enhancing drugs are actually present in the samples. When performance-enhancing drugs are present, the test is positive with probability p, and when performance-enhancing drugs are absent, the test is negative with probability q. The probability that a randomly selected player has performance-enhancing drugs in their blood is π.

(i) Find the probability that the first blood test is positive.

(ii) Find the conditional probability that performance-enhancing drugs are actually present in the sample, given that the first blood test is positive.

It is assumed the results of the two blood tests are *conditionally independent* given the presence or absence of performance-enhancing drugs in the samples.

(iii) Find the probability that the player fails the overall test.

(iv) Find the conditional probability that performance-enhancing drugs are present in the sample, given that the player fails the overall test.

3.

(a) **(i)** Suppose $X \sim \text{Binomial}(n, \theta)$, with $0 < \theta < 1$. Describe the experiment for which X is the number of 'successes.'

Small internal airlines find that each passenger who buys a seat fails to turn up with probability 1/10 independently of the other passengers. So AeroEase always sell 10 tickets for their 9-seat plane, while AirChoice always sell 12 tickets for their 10-seat planes.

(ii) Write down the probability mass functions for the numbers of people who turn up for given typical flights of AeroEase and AirChoice.

(iii) Which airline is more often over-booked?

(b) Consider the discrete random variable X with probability mass function (pmf)

$$f_X(x) = \begin{cases} 1/N, & x = 1, \dots, N \\ 0, & \text{otherwise.} \end{cases}$$

(i) Write down, in the form of a sum, the corresponding moment generating function.

(ii) The third central moment (third moment about the mean μ_X) of the discrete random variable X is defined by

$$E_{f_X}[(X - \mu_X)^3] = E_{f_X}[X^3] - 3\mu_X E_{f_X}[X^2] + 2\mu_X^3.$$

and can be used to measure the asymmetry of the distribution.

Use the moment generating function to calculate the components of the third central moment, and hence the third central moment itself.

[NB: $\sum_{x=1}^{N} x = N(N+1)/2$, $\sum_{x=1}^{N} x^2 = N(N+1)(2N+1)/6$ *and* $\sum_{x=1}^{N} x^3 = N^2(N+1)^2/4$.*]*

(iii) Given that the third central moment can be used to measure the asymmetry of the distribution, comment on your result in (ii) given the form of the pmf.

4.

A continuous random variable X has probability density function

$$f_X(x) = 4x\, e^{-2x}, \qquad 0 < x < \infty.$$

(a) Demonstrate that this pdf integrates to unity over $0 < x < \infty$, as required.

(b) Show that the cumulative distribution function is given by

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - 2x\, e^{-2x} - e^{-2x} & 0 < x < \infty. \end{cases}$$

(c) Show that the moment generating function is given by

$$M_X(t) = \frac{4}{(2-t)^2},$$

for $t < 2$.

(d) Hence find the mean, $E_{f_X}[X]$, and variance, $\mathrm{var}_{f_X}[X]$.

5.

(a) A random variable X with a normal distribution with mean μ and variance σ^2 has a moment generating function $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Let X_1 and X_2 be independent *non-identically distributed* normal random variables with means 1 and 3 and variances 2 and 4, respectively, and let $Y = X_1 + X_2$.

(i) Derive the moment generating function of Y and identify the distribution of Y.

Now let X_1 and X_2 be independent and *identically distributed* normal random variables each having a mean of 1 and variance of 2, and let $Y = X_1 + X_2$.

(ii) Derive the moment generating function of Y and identify the distribution of Y.

(iii) What do parts (i) and (ii) tell you about the distribution of the sum of normal random variables?

(b) If X_1 and X_2 are independent continuous random variables with $f_{X_1}(x_1) = 0, x_1 \leq 0$ and $f_{X_2}(x_2) = 0, x_2 \leq 0$, then the probability density function of $Y = X_1 + X_2$ is found as the convolution of the pdfs of X_1 and X_2: $f_Y(y) = \int_0^y f_{X_1}(x_1) f_{X_2}(y - x_1)dx_1$. [Note: the range of x_1 is 0 to y since $x_1 + x_2 = y$ so that x_1 cannot exceed y because x_1 and x_2 are both positive.]

An exponentially distributed random variable X has a pdf $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$.

(i) Let X_1, X_2 be *non-identically distributed* independent exponential random variables with $\lambda_1 = 2$ and $\lambda_2 = 1$, respectively. Use the convolution theorem to find the probability density function $f_Y(y)$ of $Y = X_1 + X_2$ and show that $f_Y(y) > 0$ for $y > 0$ and that $f_Y(y)$ integrates to unity.

(ii) Now let X_1, X_2 be independent *and identically distributed* exponential random variables with $\lambda_1 = \lambda_2 = 1$. Use the convolution theorem to find the probability density function $f_Y(y)$ of $Y = X_1 + X_2$.

(iii) Is the distribution of the sum in part (i) or in part (ii) that of a gamma random variable?

