THE QUEEN'S UNIVERSITY OF BELFAST

110SOR201

Level 2 Examination

Statistics and Operational Research 201

Probability and Distribution Theory

Thursday 17 January 2002 9.30 am — 12.30 pm

Examiners Professor R M Loynes
and the internal examiners

Answer **FIVE** questions

All questions carry equal marks

Approximate marks for parts of questions are shown in brackets

Write on both sides of the answer paper

1. (i) State a set of fundamental axioms concerning the probability measure P in a probability space $(\mathcal{S}, \mathcal{F}, \mathrm{P})$. **[3]**

Given two events $A, B \in \mathcal{F}$, deduce *from the axioms* that:

(a) $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$; **[2]**

(b) $\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B) = \mathrm{P}(\overline{A \cup B}) - \mathrm{P}(\overline{A})\mathrm{P}(\overline{B})$.
(*Hint:* $\overline{B} = (A \cap \overline{B}) \cup (\overline{A} \cap \overline{B})$.) **[4]**

(ii) Write down (without proof) an expression for $\mathrm{P}(A_1 \cup \cdots \cup A_n)$ in terms of the probabilities of the n events $A_1, ..., A_n \in \mathcal{F}$ and their intersections. **[2]**

A party is attended by n husband-wife pairs. For a certain game, the husbands stand in a straight line; and the wives, in random order, stand in a parallel line, so that each wife is paired with one of the husbands. Explaining each step carefully, obtain an expression for the probability that no wife is paired with her own husband. Deduce a convenient approximation to this probability when n is large. **[9]**

2. (i) Given a probability space $(S, \mathcal{F}, \mathrm{P})$ and an event $B \in \mathcal{F}$ with $\mathrm{P}(B) > 0$, define the conditional probability

$$\mathrm{Q}(A) = \mathrm{P}(A|B), \qquad A \in \mathcal{F},$$

and show that $(\mathcal{S}, \mathcal{F}, \mathrm{Q})$ is also a probability space. **[4]**

Write down (without proof) an expression for $\mathrm{P}(A_1 \cap \cdots \cap A_n)$ in terms of conditional probabilities (of the individual events $A_1, \ldots, A_n \in \mathcal{F}$), which is valid provided the events satisfy a certain condition (to be stated). **[2]**

(ii) Suppose that N dice are selected from a set of 3 fair dice in such a way that

$$\mathrm{P}(N = 1) = \mathrm{P}(N = 3) = \tfrac{1}{4}, \quad \mathrm{P}(N = 2) = \tfrac{1}{2},$$

and are then thrown. If S denotes the sum of the scores obtained, find the probability that

(a) $S = 5$, given that $N = 2$;

(b) $N = 2$, given that $S = 5$;

(c) $S = 5$, given that N is odd. **[7]**

(iii) In a series of independent games, a player has probabilities $\frac{1}{3}, \frac{5}{12}, \frac{1}{4}$ of scoring $0, 1, 2$ points respectively in any game. The series ends when the player scores 0 in a game, and the scores in individual games are then added to give a total score. Let p_n denote the probability that this total score is exactly n points. Obtain a recurrence relation for p_n, and give the values of p_0, p_1. Show, by induction or otherwise, that

$$p_n = \tfrac{3}{13}\left(\tfrac{3}{4}\right)^n + \tfrac{4}{39}\left(-\tfrac{1}{3}\right)^n, \qquad n \geq 0.$$ **[7]**

3. (i) The discrete random variables X, Y are independent, and each has the geometric distribution with parameter p, i.e.

$$\mathrm{P}(X = k) = \mathrm{P}(Y = k) = pq^{k-1}, \qquad k = 1, 2, \ldots; \quad p + q = 1.$$

(a) Determine the distribution of the random variable $Z = X + Y$. **[3]**

(b) Show that

$$\mathrm{P}(X > n) = \mathrm{P}(Y > n) = q^n, \qquad n = 0, 1, \ldots$$

and deduce that the distribution of the random variable $U = \min(X, Y)$ is geometric with parameter $p(2 - p)$. **[5]**

(ii) Let (X, Y) be discrete random variables with joint probability function $\{\mathrm{P}(X = x, Y = y) : x = x_1, x_2, ...; y = y_1, y_2, ...\}$. Define $\mathrm{E}(X|Y = y_j)$, and prove that

$$\mathrm{E}[\mathrm{E}(X|Y)] = \mathrm{E}(X).$$ **[5]**

(iii) A fair die is rolled repeatedly.

(a) Introducing indicator random variables

$$I_i = \begin{cases} 1 & \text{if face } i \text{ turns up at least once,} \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, 6$$

show that the expected number of distinct faces turning up in the course of n rolls is $6[1 - \left(\frac{5}{6}\right)^n]$. **[4]**

(b) Suppose that rolling is continued until either a five or a six turns up. Let X denote the sum of the squares of the individual scores obtained. By conditioning on the score Y obtained in the first roll, or otherwise, find $\mathrm{E}(X)$. **[3]**

4. (i) Define the P generating function (PGF) $G_X(s)$ of a count random variable X. If X is binomially distributed with parameters (n, p), show that

$$G_X(s) = (q + ps)^n \qquad (q = 1 - p),$$

and hence obtain E(X) and Var(X). Show also that, if X has the Poisson distribution

$$\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, ...; \quad \lambda > 0,$$

then

$$G_X(s) = e^{\lambda(s-1)}.$$ **[6]**

(ii) Given a sequence $X_1, X_2, ...$ of independent, identically distributed (i.i.d.) random variables, each with PGF $G(s)$, write down (without proof) expressions for the PGFs of the random variables $\sum_{i=1}^{n} X_i$ (where n is known) and $\sum_{i=1}^{N} X_i$ (where N is a further independent count random variable with PGF $G_N(s)$).

Suppose the number of car accidents in a year is Poisson distributed with parameter λ and the probabilities of an accident involving 1, 2, 3, 4 cars are 0.38, 0.55, 0.05, 0.02 respectively. Obtain an expression for the PGF of X, the total number of cars involved in accidents during a year, and derive from it the expected value of X. **[7]**

(iii) In a simple branching process, the family sizes are independent and identically distributed random variables, each with mean μ and PGF $G(s)$; X_n denotes the size of the n^{th} generation, and the initial population X_0 is 1. Explain why $G_n(s)$, the PGF of X_n, satisfies the recurrence relation

$$G_n(s) = G_{n-1}(G(s)), \qquad n \geq 1,$$

and deduce that

$$\mathrm{E}(X_n) = \mu \mathrm{E}(X_{n-1}) = \mu^n.$$

State (without proof) how the probability of ultimate extinction

$$e = \lim_{n\to\infty} \mathrm{P}(X_n = 0)$$

can be derived from $G(s)$. Determine e in the case where the family size distribution is binomial with parameters $n = 2, p = \frac{2}{3}$. **[7]**

5. (i) If $X_0, X_1, \ldots$ is a sequence of random variables defined on a state space $\{0, 1, \ldots\}$, explain what is meant by saying that $\{X_n : n = 0, 1, \ldots\}$ is a *homogeneous Markov chain* with transition probability matrix $\mathbf{P}$. If $\mathbf{p}^{(r)}$ denotes the row vector $(\mathrm{P}(X_r = 0), \mathrm{P}(X_r = 1), \ldots)$, show that

$$\mathbf{p}^{(n)} = \mathbf{p}^{(0)}\mathbf{P}^n,$$

and explain the significance of $p_{ij}^{(n)}$, the (i, j) element of $\mathbf{P}^n$. **[6]**

(ii) A homogeneous Markov chain $\{X_n : n = 0, 1, \ldots\}$ has states $\{0, 1, 2\}$ and transition probability matrix

$$\mathbf{P} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

At time $n = 0$, the system is equally likely to be in states 0, 1 or 2.

(a) Find $\mathrm{P}(X_2 = 0)$. **[2]**

(b) Explain briefly why we can be sure that a limiting distribution $\boldsymbol{\pi}$ exists, and determine it. **[6]**

(iii) A finite homogeneous Markov chain has a set T of transient states and a set A of absorbing states, and its transition P matrix is $\mathbf{P} = (p_{ij})$. When starting from state $i \in T$, let f_{ik} denote the probability of eventual absorption in state $k \in A$ and μ_i the expected time to absorption in *any* absorbing state.

Write down (without proof) sets of linear equations for the $\{f_{ik} : i \in T\}$ and the $\{\mu_i : i \in T\}$. For a chain with states $\{0, 1, 2, 3\}$ and

$$\mathbf{P} = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

determine f_{02} and μ_0. **[6]**

6. (i) The lifetime X of a certain device has c.d.f.

$$F(x) = 1 - e^{-\lambda x^2}, \qquad x \geq 0, \quad \lambda > 0.$$

Derive the p.d.f. of X, $f(x)$, and determine its mean, variance and mode.
Also determine the hazard rate function $r(x) = f(x)/[1 - F(x)]$, and briefly explain its significance.
(*Note*: the function

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \qquad p > 0,$$

has the properties

$$\Gamma(p+1) = p\Gamma(p) : \quad \Gamma(1/2) = \sqrt{\pi} : \quad \Gamma(n+1) = n!, \quad n \text{ integer} \geq 0.)$$ **[10]**

(ii) A continuous random variable X, defined over $(-\infty, \infty)$, has c.d.f. $F_X(x)$ and p.d.f. $f_X(x)$. If $Y = X^2$, describe briefly two methods whereby the p.d.f. $f_Y(y)$ may be derived. **[2]**

Using either method,

(a) find $f_Y(y)$ when

$$f_X(x) = \begin{cases} \frac{1}{9}(1+x)^2, & -1 \leq x \leq 2 \\ 0, & \text{otherwise;} \end{cases}$$ **[4]**

(b) show that, if $X \sim N(0, 1)$, i.e.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty,$$

then $Y \sim \chi^2(1)$. What is the connection between $N(0, 1)$ and $\chi^2(r)$ for $r \geq 2$?
(*Note*: the p.d.f. for the $\chi^2(r)$ distribution is

$$f_V(v) = \frac{1}{2^{r/2}\Gamma(r/2)} v^{r/2-1} e^{-v/2}, \qquad 0 \leq v < \infty, r \text{ a positive integer.})$$ **[4]**

7. (i) The continuous random variables X and Y are independent and Gamma distributed with positive parameters (α, λ) and (β, λ) respectively, i.e.

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise,} \end{cases}$$

(where the Gamma function is defined in Question 6(i)), with a similar expression for $f_Y(y)$. Show that the random variables

$$U = X + Y, \quad V = \frac{Y}{X+Y}$$

are independent, and that U is Gamma distributed with parameters $(\alpha + \beta, \lambda)$. What is the distribution of V? **[8]**

(ii) Let $X_{(1)}, ..., X_{(n)}$ be the order statistic random variables associated with random samples of size n from a distribution with p.d.f. $f(x)$ and c.d.f. $F(x)$, $-\infty < x < \infty$.

(a) Show that the p.d.f. of the largest observation, $X_{(n)}$, is

$$f_{(n)}(x) = n\{F(x)\}^{n-1} f(x), \quad -\infty < x < \infty.$$

Hence determine $\mathrm{E}(X_{(n)})$ and $\mathrm{Var}(X_{(n)})$ in the case of sampling from a uniform distribution with

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$ **[7]**

(b) It may be shown that the joint p.d.f. of $X_{(1)}, X_{(n)}$ is

$$f_{(1),(n)}(x, y) = n(n-1)\{F(y) - F(x)\}^{n-2} f(x) f(y), \quad -\infty < x < y < \infty.$$

By introducing a suitable transformation, show that, when sampling is from the uniform distribution in (a), the sample range R has the p.d.f.

$$f_R(r) = n(n-1) r^{n-2}(1-r), \quad 0 \leq r \leq 1.$$ **[5]**

8. (i) Define the moment generating function (MGF) $M_X(\theta)$ of a continuous random variable X in terms of its p.d.f. $f_X(x), -\infty < x < \infty$, and state how the MGF of $Y = a+bX$ is related to M_X. Indicate concisely two methods whereby moments of X about the origin can be derived from $M_X(\theta)$. How can these procedures be modified so as to yield (directly) moments about the mean $\mathrm{E}(X)$? **[5]**

(ii) If $Z \sim N(0,1)$, show that

$$M_Z(\theta) = \exp(\tfrac{1}{2}\theta^2),$$

and hence obtain $M_X(\theta)$, where $X = \mu + \sigma Z$. Use $M_X(\theta)$ to verify that $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. **[7]**

(iii) Show that if the distribution of X is negative exponential with parameter λ, i.e.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise,} \end{cases}$$

then

$$M_X(\theta) = \frac{\lambda}{\lambda - \theta}, \qquad \theta < \lambda.$$

If $X_i \ (i = 1, \ldots, n)$ are independent random variables, the distribution of each being negative exponential with parameter λ, and

$$S_n = \sum_{i=1}^{n} X_i,$$

what is the MGF of S_n? Show that

$$Z_n = \frac{S_n - n/\lambda}{\sqrt{n}/\lambda}$$

is asymptotically normally distributed with mean 0 and variance 1.

(*Hint*:

$$\lim_{n\to\infty} \left(1 + \frac{a}{n} + o(n)\right)^n = e^a \text{ for fixed } a. \qquad)$$

[8]

9. (i) State the assumptions which characterize a counting process $\{N(t), t \geq 0\}$ as a *Poisson process* with rate λ. Show from the assumptions that the probabilities

$$p_n(t) = \mathrm{P}[N(t) = n], \qquad n = 0, 1, \ldots$$

satisfy the equations

$$\begin{aligned} \frac{dp_0(t)}{dt} &= -\lambda p_0(t) \\ \frac{dp_n(t)}{dt} &= \lambda p_{n-1}(t) - \lambda p_n(t), \qquad n = 1, 2, \ldots \end{aligned}$$

Indicate *briefly* how these equations can be solved and quote the resulting expression for $p_n(t)$. Also state (without proof) the distribution of the inter-event times

$$X_n = T_n - T_{n-1},$$

where

$$T_n = \inf\{t : N(t) = n\}.$$ **[11]**

(ii) It can be shown that, for a 'birth and death' process $\{X(t),\ t \geq 0\}$ with 'birth' rates $\{\alpha_i;\ i = 0, 1, \ldots\}$ and 'death' rates $\{\beta_i;\ i = 1, 2, \ldots\}$, the probabilities

$$p_n(t) = \mathrm{P}[X(t) = n], \qquad n = 0, 1, \ldots$$

satisfy the equations

$$\frac{dp_n(t)}{dt} = -(\alpha_n + \beta_n)p_n(t) + \alpha_{n-1}p_{n-1}(t) + \beta_{n+1}p_{n+1}(t), \quad n = 0, 1, \ldots,$$

where $\alpha_{-1} = \beta_0 = 0$. Show that, if a steady state distribution $\{\pi_m;\ m = 0, 1, \ldots\}$ exists, then

$$\alpha_m \pi_m = \beta_{m+1}\pi_{m+1}, \qquad m = 0, 1, \ldots,$$

and deduce that

$$\pi_0 = \left(1 + \frac{\alpha_0}{\beta_1} + \frac{\alpha_0\alpha_1}{\beta_1\beta_2} + \cdots\right)^{-1}.$$ **[5]**

(iii) Consider a single-server queueing system in which the service time is negative exponential with mean μ^{-1} and customer arrivals form a Poisson process with rate λ, except that any customer arriving when there are already N customers in the system leaves without joining the queue. Show that the steady-state distribution of the number of customers in the system is

$$\pi_n = \rho^n(1-\rho)(1-\rho^{N+1})^{-1}, \quad 0 \leq n \leq N,$$

where $\rho = \lambda/\mu$. **[4]**