UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Thursday 5 June 2008 9.00 to 11.00

INTRODUCTION TO PROBABILITY

*There are **FIVE** questions in total.*

*Attempt **both** questions in **Section A***
*and at most **one** question in **Section B**.*

Each question in Section A is worth 30 marks
and each question in Section B is worth 40 marks.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

SECTION A

1 Let $Y_1, Y_2, \ldots$ be a sequence of random variables and let Y be a random variable. Define what it means to say that $Y_n \to Y$ in distribution as $n \to \infty$. Assume that $X_1, X_2, \ldots$ are independent identically distributed exponential random variables with parameter $\lambda > 0$, and let

$$M_n = \max_{1 \le m \le n} X_m.$$

Show that

$$M_n - \frac{1}{\lambda} \log(n) \underset{n\to\infty}{\longrightarrow} M$$

in distribution, where the cumulative distribution function of M is given by

$$P(M \le y) = \exp(-e^{-\lambda y}), \quad \text{for all } y \in \mathbb{R}.$$

What is the density of the random variable M ? (The distribution of M is known as the Gumbel distribution).

2 (a) Let $(S_n, n \ge 0)$ be an irreducible Markov chain on an infinite state space S. Define what it means for $(S_n, n \ge 0)$ to be transient, and give two necessary and sufficient conditions for this to occur. Briefly describe what happens when these conditions are not satisfied.

(b) Let $(X_n, n \ge 1)$ be independent identically distributed random variables, such that $P(X_n = +1) = p > 1/2$ and $P(X_n = -1) = 1 - p$. Let $(S_n, n \ge 0)$ be the Markov chain defined by $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$. Show that $\lim_{n\to\infty} S_n/n = c$ exists and $c \ne 0$. Deduce that $(S_n, n \geqslant 0)$ is transient.

(c) Let $x = P_0(T < \infty)$ where $T = T_{-1} := \inf\{n \ge 1 : S_n = -1\}$ and where S_n is as in (b). By conditioning on the first step, show that x satisfies a certain equation. Conclude that $x = (1-p)/p$.

SECTION B

3 Let $(X_n, n \in \mathbb{Z})$ be a sequence of independent identically distributed random variables such that $P(X_i = +1) = 1/2 = P(X_i = -1)$. For $n \geq 0$, let

$$\ell_n = \begin{cases} 0 & \text{if } X_n = -1 \\ \max\{m \geq 1 : X_{n-m+1} = \ldots = X_n = 1\} & \text{if } X_n = 1. \end{cases}$$

[If we think of coin-tossing (with $X_i = +1$ standing for a head and $X_i = -1$ for a tail), then ℓ_n is the length of the run of heads at time n, i.e., the number of successive heads between time n and the last time there was a tail.] Let

$$L_n = \max_{1 \leq m \leq n} \ell_m$$

be the longest run at time n.

(a) Assume $X_0 = -1$, $X_1 = 1$, $X_2 = 1$, $X_3 = -1$, $X_4 = 1$ and $X_5 = -1$, so that $\ell_1 = 1$, $\ell_2 = 2$, $\ell_3 = 0$, $\ell_4 = 1$ and $\ell_5 = 0$. What are the values of $L_1, \ldots, L_5$?

(b) Define almost sure convergence and draw a diagram which shows how different types of convergence imply other types of convergence. State the Borel–Cantelli lemmas, assuming definitions from the lectures.

(c) Let $n \geqslant 1$ and let $\log_2 n := (\ln n)/(\ln 2)$. (In particular, note that $2^{\log_2 n} = n$.) Find the distribution of ℓ_n, and deduce that almost surely for any $\epsilon > 0$, $\ell_n \leq (1+\epsilon)\log_2 n$ for large enough (say $n \geq N(\epsilon, \omega)$). Deduce that

$$\limsup_{n \to \infty} \frac{L_n}{\log_2 n} \leq 1\,.$$

4 We consider the following model for the evolution of the size of a population $(Z_n, n \geq 0)$. At time $n = 0$, the population consists of just one individual, so $Z_0 = 1$. The Z_n are then constructed inductively as follows: given Z_n,

$$Z_{n+1} = \begin{cases} 0 & \text{if } Z_n = 0 \\ \sum_{i=1}^{Z_n} X_i^{(n)} & \text{if } Z_n > 0 \end{cases}$$

where the random variables $(X_i^{(n)}, i \geq 1, n \geq 0)$ are independent identically distributed Poisson random variables with parameter λ. That is, each individual in the population at time n has an independent Poisson number of offspring, which make up the population at time $n+1$. In this question, results from lectures may be used without proof but should be stated clearly.

(a) Give the definition of a martingale.

(b) Show that $(M_n = Z_n/\lambda^n, n \geq 0)$ defines a martingale. Conclude when $\lambda < 1$ that $E(Z_n) \to 0$ and then that $Z_n \to 0$ almost surely as $n \to \infty$.

(c) Let $\lambda > 1$. By conditioning on $\{Z_n = k\}$ and using the fact that $E(X^2) = \lambda + \lambda^2$ when $X \sim \text{Poisson}(\lambda)$, show that

$$E(M_{n+1}^2) = E(M_n^2) + \frac{1}{\lambda^{n+1}}.$$

Using a martingale convergence result, deduce that M_n converges in L^2 to some limit M. Show that $E(M) = 1$ and deduce that with positive probability, $Z_n > 0$ for all n.

5 Let $N \geq 1$ and consider the following model for the evolution of the genetic material of a population whose size is constant equal to N. In this model, individuals can be of only two types, a and A. The population at generation $n+1$ is obtained from the population at generation n as follows. For every $1 \leq i \leq N$, individual i chooses a parent uniformly at random among individuals of generation n, independently of all other individuals, and adopts her type.

[A formal definition of the model follows. For $1 \leq i \leq N$, and $n \geq 0$, let $Y_i(n) \in \{a, A\}$ be the type of individual i at generation n. Let $(U_i(n), 1 \leq i \leq N, n \geq 0)$ be a family of independent identically distributed uniform random variables on $\{1, \ldots, N\}$, then

$$Y_i(n+1) = Y_{U_i(n)}(n), \quad 1 \leq i \leq N\,.$$

Let X_n be the number of individuals of type a in generation n. We assume that initially $X_0 = x \in \{1, \ldots, N\}$.]

In this question, results from lectures may be used without proof but should be stated clearly.

(a) Show that $(X_n, n \geq 0)$ is a Markov chain on $\{0, \ldots, N\}$ with transition probabilities:

$$p(i,j) = \binom{n}{j} \left(\frac{i}{n}\right)^j \left(1 - \frac{i}{n}\right)^{n-j}$$

and that $(X_n, n \geq 0)$ is a martingale.

(b) Show that, no matter what the starting point, ultimately, $X_n = 0$ or $X_n = N$. (In the latter case, the a subpopulation has invaded the population). If $T_i = \inf\{n \geq 0 : X_n = i\}$, show that

$$P(T_N \leq T_0 | X_0 = x) = x/N$$

(c) We now make the following change to the model to account for the possibility of mutations. Assume that individuals choose their parent at random as before, but adopt the type of their parent only with probability $0 < \mu < 1$, while they adopt the opposite type with probability $1-\mu$. We assume that occurrence of mutations is independent for different individuals and generations. Show that, in that case, X_n converges in distribution to some non-trivial variable X independent of the starting point, and such that $P(X = i) > 0$ for all $0 \leq i \leq N$. (You are not asked to compute explicitly this distribution).

END OF PAPER