UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Thursday, 27 May, 2010 1:30 pm to 3:30 pm

INTRODUCTION TO PROBABILITY

*Attempt no more than **THREE** questions.*

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

The questions carry equal weight.

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.

1

(a) Let $(X_n,\, n \geqslant 0)$ be a sequence of random variables, and let X be a random variable. What does it mean to say that X_n converges to X in distribution? Show that if X_n and X take values in $\{0,\, 1,\, \ldots\}$, then this is equivalent to

$$\mathbb{P}\,(X_n = k) \;\to\; \mathbb{P}\,(X = k)$$

as $n \to \infty$, for every $k \in \{0,\, 1,\, \ldots\}$. Let c_n be a sequence of nonnegative numbers such that $c_n \to c > 0$ as $n \to \infty$. Let X_n have a binomial distribution with parameters $(n,\, c_n/n)$. Show that X_n converges to X in distribution, for some random variable X to be determined.

(b) Let $\lambda > 0$, and suppose that $N = \lfloor \lambda n \rfloor$ balls are placed uniformly at random, independently from one another, in n urns labelled 1 through n. Let $Z_n(i)$ denote the total number of balls in urn number i when all balls have been placed. What is the distribution of $Z_n(i)$ for a fixed $1 \leqslant i \leqslant n$? Explain briefly why the random variables $(Z_n(i))_{1 \leqslant i \leqslant n}$ cannot be independent. Find a random variable Z such that $Z_n(i) \to Z$ in distribution as $n \to \infty$.

(c) Deduce the following: let W_n denote the number of empty urns when all balls have been placed. Then as $n \to \infty$,

$$\mathbb{E}(W_n) \;\sim\; e^{-\lambda}\, n\,,$$

i.e, the ratio of the two sides converges to 1.

2

Let X and Y be two independent Gaussian random variables with mean 0 and variance 1.

(a) By studying the characteristic functions or otherwise, show that $Z = X - Y$ is a Gaussian random variable. What is its variance? Deduce that as $\varepsilon \to 0$,

$$\mathbb{P}\left(|X - Y| < \varepsilon\right) \sim \frac{\varepsilon}{\sqrt{\pi}},$$

i.e., the ratio of the two sides tends to 1.
[*Hint: You may use without proof the following fact from calculus: if* $f : \mathbb{R} \to \mathbb{R}$ *is continuous, then for all* $x \in \mathbb{R}$, *as* $\varepsilon \to 0$, $\int_{x-\varepsilon}^{x+\varepsilon} f(s)\,ds \sim 2\varepsilon f(x)$.]

(b) Fix $x \in \mathbb{R}$, and $\varepsilon > 0$. Express $\mathbb{P}\left(X \leqslant x,\, |X - Y| < \varepsilon\right)$ as a double integral. Show that

$$\mathbb{P}\left(X \leqslant x,\, |X - Y| < \varepsilon\right) \sim \frac{\varepsilon}{\pi} \int_{-\infty}^{x} e^{-s^2}\,ds$$

as $\varepsilon \to 0$. You may take limits under the integral sign without justification.

(c) Let X_ε denote the law of X given that $|X - Y| < \varepsilon$. Deduce from parts (a) and (b) above that X_ε converges in distribution towards a certain random variable W, which you should identify.

(d) Now, consider the pair of random variables $(X,\, X - Y)$. Write down the joint density of this pair of random variables. What is the conditional density of X given $X - Y = 0$?

3

(a) State and give a proof of the Borel-Cantelli lemmas.

(b) Let $X_1, \ldots X_n$ be random variables such that $var(X_i) < \infty$ for all $1 \leqslant i \leqslant n$. Show that

$$var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} var(X_i) + 2 \sum_{1 \leqslant i < j \leqslant n} cov(X_i, X_j).$$

[*Hint: it may help to introduce the random variable* $Y_i = X_i - \mathbb{E}(X_i)$.]

Hence deduce that $var(X_1 + \ldots + X_n) < \infty$.

(c) Let $(A_1, A_2, \ldots)$ be a sequence of events such that $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$. We also assume that these events are *pairwise* independent, i.e., for every $i \neq j$, A_i and A_j are independent. For $k \geqslant 1$ define a random variable N_k by

$$N_k = \sum_{i=1}^{k} \mathbf{1}_{A_i}.$$

Let $m_k = \mathbb{E}(N_k)$. Show that $m_k \to \infty$ as $k \to \infty$ and that $var(N_k) \leqslant m_k$.

(d) Using Chebyshev's inequality, show that $\mathbb{P}(N_k \leqslant m_k/2) \to 0$ as $k \to \infty$. Conclude that the events A_i occur infinitely often almost surely. Explain briefly why this result is stronger than the second Borel-Cantelli lemma.

4

Let $S = \{1, 2, 3\}$ and consider a Markov chain $X = (X_0, X_1, \dots)$ with values in S defined by the transition matrix P defined as follows:

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Draw a diagram to represent the possible one-step transitions of X, including the transition probabilities. What happens if $X_0 = 3$? Is X irreducible?

(b) Define a stopping time τ by $\tau = \inf\{n \geqslant 0 : X_n = 3\}$. For $n \geqslant 0$ and $x \in S$ define $\alpha_n(x) = \mathbb{P}_x(\tau > n)$. Show $\alpha_n(2) = \frac{1}{3}\,\alpha_{n-1}(1)$, and deduce a recurrence relation for $\alpha_n(1)$. Deduce that there exists $A, B \in \mathbb{R}$ and $\lambda > \mu$ such that $\alpha_n(1) = A\lambda^n + B\mu^n$ (it is not asked to compute A and B but you should find the value of λ and μ). Conclude that

$$\alpha_n(1) \sim A\left(\frac{2}{3}\right)^n$$

as $n \to \infty$, where $A \in \mathbb{R}$ is the same as above.

(c) Let $S' = \{1, 2\}$. Using part (b) above, compute for $x, y \in S'$,

$$q(x, y) = \lim_{n \to \infty} \mathbb{P}_x(X_1 = y \mid \tau > n).$$

Show that, for any $k \geqslant 1$ and $x_1, \dots, x_k \in S'$, as $n \to \infty$:

$$\lim_{n \to \infty} \mathbb{P}_x(X_1 = x_1, \dots, X_k = x_k \mid \tau > n) = \mathbb{P}_x(Y_1 = x_1, \dots, Y_k = x_k)$$

where $Y = (Y_0, Y_1, \dots)$ is the Markov chain on S' with transition probabilities determined by $q(x, y)$.

(d) Determine whether or not Y has an invariant distribution, and find the invariant distribution if it exists.

5

A gambler plays at a game of coin tossing against a casino. He adopts the following strategy. He always bets on heads, and retires from the game the first time the coin comes up tails, or in any case after three rounds, whichever comes first. Each time he wins his bet, the casino pays him double his bet, and he reinvests his winnings in the next round. His initial fortune is 0 and his initial bet is 1, in pounds. Thus, at the end of the first round, his total winnings may be 2 or 0 (corresponding to a fortune of 1 or -1), at the second round his total winnings may be 4 or 0 (corresponding to a fortune of 3 or -1), and at the third round his winnings may be 8 or 0 (corresponding to a fortune of 7 or -1).

(a) Let Y be fortune of the player at the end of the game, in pounds. Assuming the coin is fair, what is the probability mass function of Y? Show that $\mathbb{E}(Y) = 0$.

(b) The casino now tosses a fair coin infinitely often and we denote by $(X_1, X_2, \dots)$ the successive outcomes. Before each time $i = 1, 2, \dots,$ a new gambler arrives and plays the game described above part (a), using the outcomes (X_i, X_{i+1}, X_{i+2}). Note that outcome X_i may be relevant to gamblers i, $i-1$, $i-2$ if they have not retired already. We are interested in the time τ it takes for one of the gamblers to win, i.e., for the pattern HHH to occur in this sequence.

Let M_n denote the cumulative fortune of gamblers 1 through n at time n, and let G_n denote the cumulative winnings of players 1 though n by time n. Show that $M_n = -n + G_n$. Explain why $(M_1, M_2, \dots)$ is a martingale with respect to the filtration generated by $(X_1, X_2, \dots)$.

[It may help to note that each gambler's bet is fair no matter how much they are gambling.]

(c) Deduce from part (b) that for all $n \geqslant 1$, $\mathbb{E}(M_{n\wedge\tau}) = 0$. [Any application of a theorem from the course must be properly justified.] Show that $G_\tau = 14$ and that $|G_n| \leqslant 14$ for all $n \leqslant \tau$. Deduce that $\mathbb{E}(\tau) = 14$.

(d) Suppose now that the pattern we are interested in is HTH. Let τ' be the time this pattern first occurs in the sequence of coin tosses. By changing the strategy of each gambler so that they can only win if the pattern HTH occurs, show using the same method that now $\mathbb{E}(\tau') = 10$. (Non-examinable: does this surprise you?)

END OF PAPER