MATHEMATICAL TRIPOS: PART IA **Lent 2022**

PROBABILITY **Dominic Yeo**

Example Sheet 3 (of 4)

1. Let X be a random variable.

(a) Show that, for all $p \in (0, \infty)$ and all $x \in (0, \infty)$,

$$\mathbb{P}(|X| \geq x) \leq \mathbb{E}(|X|^p)x^{-p}.$$

(b) Show that, for all $z \geq 1$,

$$\mathbb{P}(X \geq x) \leq \mathbb{E}(z^X)z^{-x}.$$

2. Let X_n be a binomial random variable corresponding to n independent trials, each with success probability 1/3. Find upper bounds on $\mathbb{P}(X_n \geq \frac{2}{3}n)$ using (a) Markov's inequality; (b) Chebyshev's inequality; (c) the *Chernoff bound* method suggested in Problem 1(b). Comment on the quality of these bounds when n is large.

3. For a random variable X with mean μ and variance $\sigma^2 < \infty$, define the function

$$V(x) = \mathbb{E}[(X - x)^2].$$

Express the random variable $V(X)$ in terms of μ, σ^2 and X, and hence show that

$$\mathbb{E}[V(X)] = 2\sigma^2.$$

[*Hint: Re-characterise V before trying to 'substitute in' the random variable X.*]

4. Consider a random sample $X_1, \ldots, X_n$ taken from a distribution having mean μ and variance $\sigma^2 < \infty$. Use Chebyshev's inequality to determine a sample size n that will be sufficient, whatever the distribution, for the probability to be at least 0.99 that the sample mean $\bar{X}$ will be within two standard deviations of μ.

5. Let $(X_n : n \in \mathbb{N})$ be a sequence of independent identically distributed random variables, with mean μ and variance $\sigma^2 < \infty$. Set $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Let N be a bounded non-negative integer-valued random variable which is independent of the sequence $(X_n : n \in \mathbb{N})$.

(a) Recall the proof from lectures that $\mathbb{E}[S_N] = \mu\mathbb{E}[N]$.

Show that $\mathbb{E}(S_N^2 | N = n) = n\sigma^2 + n^2\mu^2$ and hence express $\mathrm{var}(S_N)$ in terms of $\mathrm{var}(N)$.

(b) Verify that $\mathbb{E}[S_N] = \mu\mathbb{E}[N]$ by differentiating the PGF of S_N.

(c) Find a counterexample to the result $\mathbb{E}[S_N] = \mu\mathbb{E}[N]$ when N is not independent of (X_n).

6. A bug jumps around the vertices of a triangle, labelled $\{1,2,3\}$. At every jump, it moves from its current position to one of the other two vertices with probability $1/2$ each (independently of how it arrived at its current position). The bug starts at vertex 1.

(a) Let p_n be the probability that the bug is at vertex 1 after n jumps. Find p_n for each $n \geq 0$.

(b) What happens to p_n as $n \to \infty$?

(c) Let N be the number of jumps until the bug visits vertex 3. Find $\mathbb{E}[N]$.

[*The conclusion of part (b) is extended with great generality in Part 1B Markov Chains.*]

7. A slot machine operates so that at the first turn the probability for the player to win is $1/2$. Thereafter the probability for the player to win is $1/2$ if he lost at the last turn, but is $p < 1/2$ if he won at the last turn. If u_n is the probability that the player wins at the nth turn, show that, provided $n > 1$,

$$u_n + (\tfrac{1}{2} - p)u_{n-1} = \tfrac{1}{2}.$$

Observe that this equation also holds for $n = 1$, if we set $u_0 = 0$. Solve the equation, showing that

$$u_n = \frac{1 + (-1)^{n-1}(\frac{1}{2} - p)^n}{3 - 2p}.$$

8. Suppose we conduct a sequence of independent Bernoulli trials and denote by X the number of trials up to and including the ath success.

(a) Show that

$$\mathbb{P}(X = r) = \binom{r-1}{a-1} p^a q^{r-a}, \quad r = a, a+1, \ldots.$$

(b) Show that the generating function G_X for this distribution is $p^a t^a (1 - qt)^{-a}$.

Deduce that $\mathbb{E}(X) = a/p$ and $\text{var}(X) = aq/p^2$.

[*Hint: consider writing X as a sum of independent random variables.*]

(c) Find an expression for $\mathbb{P}(X \text{ is even})$.

9. Let X have the uniform distribution on $\{1,2,\ldots,6\}$, and let X_1, X_2 be two independent copies of X, representing rolling a dice twice. Suppose Y is any probability distribution on $\{1,2,\ldots,6\}$, with Y_1, Y_2 two independent copies of Y.

(a) Find the generating function of $X_1 + X_2$.

(b) Suppose that $X_1 + X_2$ has the same distribution on $\{2,3,\ldots,12\}$ as $Y_1 + Y_2$. Prove that Y must itself be uniform.

(c) Now suppose that Y has the uniform distribution on $(y_1, y_2, \ldots, y_6)$ where the y_i are (not necessarily distinct) integers between 1 and 11. Suppose Z has the uniform distribution $(z_1, z_2, \ldots, z_6)$ satisfying the same conditions. If Y, Z are independent, and $Y + Z$ has the same distribution as $X_1 + X_2$, does this imply that $\{y_1, \ldots, y_6\} = \{z_1, \ldots, z_6\} = \{1, \ldots, 6\}$?

10. At time 0, a blood culture starts with one red cell. At the end of one minute, the red cell dies and is replaced by one of the following combinations with probabilities as indicated:

$$2 \text{ red cells } \tfrac{1}{4}, \quad 1 \text{ red, } 1 \text{ white } \tfrac{2}{3}, \quad 2 \text{ white } \tfrac{1}{12}.$$

Each red cell lives for one minute and gives birth to offspring in the same way as the parent cell. Each white cell lives for one minute and dies without reproducing. Assume the individual cells behave independently.

(a) At time $n + \frac{1}{2}$ minutes after the culture began, what is the probability that no white cells have yet appeared?

(b) What is the probability that the entire culture dies out eventually?

11. Consider a population of animals in which each mature individual produces a random number of offspring with generating function F. Suppose we start with a population of k immature individuals, each of which grows to maturity with probability p, independently of the other individuals.

(a) Find the generating function for the distribution of the number of immature individuals in the next generation.

(b) Find the generating function for the distribution of the number of mature individuals in the next generation, given that there are k mature individuals in the parent generation.

(c) Show that the distributions in (a) and (b) have the same mean, but not necessarily the same variance.

12. Let $F(t) = 1 - p(1-t)^\beta$, where $p \in (0,1)$ and $\beta \in (0,1)$ are constants. Show that $F(t)$ is the generating function of a probability distribution on $\mathbb{Z}^+$ and that its iterates are given by

$$F_n(t) = 1 - p^{1+\beta+\cdots+\beta^{n-1}}(1-t)^{\beta^n} \quad \text{for } n = 1, 2, \ldots.$$

Find the mean m of the associated distribution and the extinction probability of the branching process whose offspring distribution has generating function F.

Extensions

13. Let q_λ be the extinction probability of a branching process with Poisson(λ) offspring distribution. Consider $\zeta_\lambda = 1 - q_\lambda$ the corresponding *survival probability*. Show that ζ_λ satisfies $\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}$. Describe the behaviour of $\zeta_{1+\epsilon}$ as $\epsilon \downarrow 0$, in the form $\zeta_{1+\epsilon} \sim C\epsilon^\alpha$, for constants C, α to be determined.

[*Hint: start by showing* $\zeta_\lambda \to 0$ *as* $\lambda \downarrow 1$, *then expand.*]

14. Let G be a *graph*, consisting of a finite collection of vertices $V(G)$, some pairs of which are connected by an edge. We declare the *neighbours* of a vertex v to be those vertices w such that v, w are directly connected by an edge. We assume the graph is *connected*, so that there is a path of edges joining any pair of vertices. Let A be a non-empty subset of the vertices, and $B = V(G) \setminus A$. For each vertex $v \in A$, we declare a value $a_V \in \mathbb{R}$.

Use a probabilistic argument to show that there exists a function $f : V(G) \to \mathbb{R}$ such that $f(v) = a_v$ for all $v \in A$, and for all $v \in B$, $f(v)$ is the average of the values taken by f on the neighbours of v.

Can you justify that f is unique?

[*Note: such a function f is called the* (discrete) harmonic extension *of* (a_v)*.*]

15. Let $\mathcal{T}$ be a branching process tree with offspring distribution X satisfying $\mu = \mathbb{E}[X] \leq 1$. Order the individuals in a *breadth-first* manner, so that the root is x_1, and the the children of the root are $x_2, \ldots, x_{1+Z_1}$, and the individuals in the $(n+1)$th generation are $x_{Z_0+Z_1+\ldots+Z_n+1}, \ldots, x_{Z_0+Z_1+\ldots+Z_n+Z_{n+1}}$. Let $c(x_i)$ be the number of children of individual x_i.

(a) Consider the random process given by $S_0 = 0$ and $S_m = c(x_1) + \ldots + c(x_m) - m$. Explain briefly why $(S_0, S_1, \ldots)$ is a random walk whose increments have distribution $X - 1$.

(b) Show that $|\mathcal{T}|$, the total number of individuals in the population, has the same distribution as $\tau := \inf\{m \geq 0 : S_m = -1\}$.

(c) Prove that $\mathbb{P}(|\mathcal{T}| = m) = \frac{1}{m}\mathbb{P}(S_m = -1)$.

(d) Suppose that X takes the values 2 and 0 each with probability 1/2. Explain why $\mathbb{E}[|\mathcal{T}|] = \infty$, and find constants C, α such that $\mathbb{P}(|\mathcal{T}| = m) \sim Cm^{-\alpha}$.

[*Hint: you may find a question on a previous sheet helpful.*]

16. Let $\mathcal{T}$ be a branching process tree with supercritical offspring distribution X satisfying $\mu = \mathbb{E}[X] > 1$. Denote by Ψ the *extinction* event $\{|\mathcal{T}| < \infty\}$, and assume $q = \mathbb{P}(\Psi) \in (0, 1)$.

(a) Explain briefly why the conditional branching process tree $(\mathcal{T} \mid \Psi)$ is itself a branching process, and describe its offspring distribution $\hat{X}$.

(b) Show that if $X \sim \mathrm{Po}(\mu)$, then $\hat{X} \sim \mathrm{Po}(\nu)$, where $\nu \neq \mu$ and satisfies $\mu e^{-\mu} = \nu e^{-\nu}$. Show that ν is monotone as a function of μ.

(c) Now return to $\mathcal{T}$, with $X \sim \mathrm{Po}(\mu)$. Colour blue all individuals with infinitely many descendents, and colour red all others, so that $\Psi^c = \{\text{root is blue}\}$. State the distribution of the number of blue children of the root. Now characterise the distribution of the number of blue children of the root, conditional on Ψ^c, and check that the conditional mass function sums to 1.

Give as complete as description as you can manage for the structure of the blue and red individuals in $\mathcal{T}$, conditional on Ψ^c.