MATHEMATICAL TRIPOS: PART IA **Lent 2022**

PROBABILITY **Dominic Yeo**

Example Sheet 4 (of 4)

1. Let $X \sim \text{Exp}(\lambda)$.

(a) Find $\mathbb{P}(a < X < b)$ for $0 < a < b$.

(b) Find $\mathbb{P}(\sin X > \frac{1}{2})$.

(c) For $x \in \mathbb{R}$, the *ceiling*, denoted $\lceil x \rceil$, of x is the smallest integer greater than or equal to x. Find the distribution of $\lceil X \rceil$.

2. The area of a circle is exponentially distributed with parameter λ. Find the probability density function of the radius of the circle. Have you seen this distribution before for some value of λ?

3. A stick is broken in two places, independently uniformly distributed along its length. What is the probability that the three pieces will make a triangle?

4. The random variables X and Y are independent and exponentially distributed with parameters λ and μ respectively. Find the distribution of $\min\{X, Y\}$ and find $\mathbb{P}(X > Y)$.

5. A radioactive source emits particles in a random direction (with all directions being equally likely). It is held at a distance a from a vertical infinite plane photographic plate.

(a) Show that, given the particle hits the plate, the horizontal coordinate of its point of impact (with the point nearest the source as origin) has the Cauchy density function $a/(\pi(a^2 + x^2))$.

(b) Can you compute the mean of this distribution?

6. A random variable X is said to have *log-normal distribution* if $Y = \log X$ is normally distributed.

(a) Find the mean and variance of X in the case where $Y \sim N(\mu, \sigma^2)$.

(b) Log-normal distributions are used to model quantities X which are believed to arise as the product of many positive random factors $X = \xi_1 \xi_2 \dots \xi_n$, such as particle sizes after a crushing process or stock prices. Making any reasonable assumptions you wish, give a justification for such a model.

7. The random variables X and Y are independent, with joint density $f_{X,Y}$.

(a) Let $Z = X + Y$. Find the joint distribution of (X, Z), and then the marginal of Z, recovering the convolution result from lectures.

(b) Assume $X, Y \sim \text{Exp}(\lambda)$. Show that the random variables $X+Y$ and $X/(X+Y)$ are independent and find their distributions.

8. Let $x_1, x_2, \ldots, x_n$ be positive real numbers.

(a) Show that their harmonic mean is no greater than their arithmetic mean, that is,

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1} \le \frac{1}{n}\sum_{i=1}^{n} x_i.$$

(b) Show that, if $y_1, \ldots, y_n$ is any reordering of $x_1, \ldots, x_n$, then

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i} \ge 1.$$

9. Suppose X is a real-valued random variable and $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are two non-decreasing functions. Prove the 'Chebyshev order inequality':

$$\mathbb{E}(f(X))\mathbb{E}(g(X)) \le \mathbb{E}(f(X)g(X)).$$

[*Hint. Consider* $(f(X_1) - f(X_2))(g(X_1) - g(X_2))$ *where* X_1 *and* X_2 *are independent copies of* X.]

10. Fix $\lambda > 0$, and for every $n \ge 1$, let $X_n \sim \mathrm{Geom}(\lambda/n)$, supported on $\{1, 2, \ldots\}$. Prove that $\frac{X_n}{n}$ converges in distribution to $\mathrm{Exp}(\lambda)$, (i) directly; (ii) using MGFs.

11. (a) Consider a random sample $(X_1, \ldots, X_n)$ taken from a normal distribution $N(\mu, \sigma^2)$. How large must n be to ensure that the probability that the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is within one standard deviation σ of the mean μ is at least 0.99?

[*Hint. For the distribution function* Φ *of* $N(0,1)$ *we have* $\Phi(2.58) = 0.995$.]

(b) A random sample is taken in order to find the proportion p of left-handed people in a large population. Guided by the central limit theorem, determine a sample size such that the probability of a sampling error less than 0.04 will be 0.99 or greater.

12. Show that, as $n \to \infty$,

$$e^{-n}\left(1 + \frac{n}{1!} + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right) \to \frac{1}{2}.$$

Extensions

13. Let us define discrete random variables X and Y by a joint probability mass function as shown:

Y \ X	-1	0	1
-1	a	b	c
0	d	e	f
1	g	h	i

Write down the marginal mass functions $p_X(\cdot)$ and $p_Y(\cdot)$ of X and Y.

A lecturer is trying to choose $\{a, b, \ldots, i\}$ to illustrate the general principle that $\mathrm{Cov}(X, Y) = 0$ does *not* imply X, Y independent. Describe geometrically the sets of $\{a, b, \ldots, i\}$s which (i) induce independence, and (ii) induce zero covariance, and conclude that if the lecturer chooses *uniformly at random from the set in (ii)* (*), this will illustrate their point with probability one.

[You may reflect on how to make (*) formal, but this is not part of the question.]

14. The Polya urn model for contagion is as follows. We start with an urn which contains one white marble and one black marble. At each second we choose a marble at random from the urn and replace it together with one more marble of the same colour. Calculate the probability that when n marble are in the urn, i of them are white, and conclude a limit result for convergence in distribution of the proportion of white marbles in the urn.

Can you strengthen this to an almost sure limit result?

[*Hint: start by simulating the limit X and consider, conditional on X, an IID sequence of biased coin tosses with appropriate parameter.*]

15. We recall Q15 from Sheet 2. Let σ_n be a uniformly chosen permutation from Σ_n, and α_n the number of cycles in σ_n. Denote by $\mathcal{C} = (C_1, C_2, \ldots, C_{\alpha_n})$ the lengths of these cycles, in decreasing order. Note that conditional on $\mathcal{C}$, the elements $\{1, 2, \ldots, n\}$ are assigned to the cycles uniformly.

Let ℓ_n be the length of the cycle containing the element 1. Let β_n be the length of a cycle chosen uniformly from the α_n possible cycles. Find expressions for $\mathbb{E}[\ell_n \,|\, \mathcal{C}]$ and $\mathbb{E}[\beta_n \,|\, \mathcal{C}]$ in terms of $\mathcal{C}$, and prove that $\mathbb{E}[\ell_n] \geq \mathbb{E}[\beta_n]$. Finally, prove that $\mathbb{E}[\beta_n]\mathbb{E}[\alpha_n] \geq n$.

16. Daniel is playing *Snakes and Ladders*, which we model by adapting a random walk.

(a) To avoid tears, initially there are no snakes. Let $X_1, X_2, \ldots$ be IID uniform choices from $\{1, 2, \ldots, 6\}$, denoting dice rolls, with $S_n = X_1 + \ldots + X_n$ denoting position after n moves.

Daniel plays until time $T_N := \inf\{n \geq 0 \,:\, S_n \geq N\}$, where N is large for the sake of his parents' productivity. Derive a CLT for T_N, ie a limit in distribution for $\frac{T_N - a_N}{b_N}$, for some a_N, b_N.

(b) Let $\alpha_N = \lfloor N/3 \rfloor$ and $\beta_N = \lfloor 2N/3 \rfloor$. We introduce a snake from $\beta_N \mapsto \alpha_N$, so that now

$$S_{n+1} := \begin{cases} S_n + X_n & \text{if } S_n + X_n \neq \beta_N \\ \alpha_N & \text{if } S_n + X_n = \beta_N. \end{cases}$$

When N is large, state the limiting probability that the second case (ie the snake move) occurs at least once. [You do not need to prove the validity of this limit.]

For large N, describe approximately the distribution of $T_N := \inf\{n \geq 0 \,:\, S_n \geq N\}$, and explain why there is no choice of a_N, b_N such that $\frac{T_N - a_N}{b_N}$ has a continuous limit in distribution.