1.1 Random Binary Expansions

This project requires an understanding of the Part IA Probability and Part IA Analysis courses.

Let $U = (U_1, U_2, \dots)$ represent an infinite sequence of coin tosses, with $U_i = 1$ if the ith toss is heads and $U_i = 0$ if it is tails. Suppose the coin tosses are independent, and that the probability of heads is $p \in (0, 1)$ and the probability of tails is $q = 1 - p$.

Given such a sequence we can define a real-valued random variable $X = f(U)$, taking values in the interval $[0, 1]$, by

$$f(U) = \sum_{i=1}^{\infty} \frac{U_i}{2^i}.$$

We may think of U as a binary expansion of X (though in fact some $x \in [0, 1]$ do not have a unique binary expansion, to wit, $0.1 = 0.01111\dots$).

Define the cumulative distribution function

$$F(x) = \mathbb{P}(X \leqslant x).$$

For most values of p the function F is pathological, but it does have some interesting properties.

Approximating F

One way to approximate F is by Monte Carlo simulation, as follows. Fix $n \in \mathbb{N}$. Generate a finite sequence $U^n = (U_1, \dots, U_n)$, and compute $X^n = \sum_{i \leqslant n} U_i/2^i$. Repeat this N times to generate a random sample $X_1^n, \dots, X_N^n$. Now we can plot the *empirical* cumulative distribution function

$$\hat{F}(x) = \frac{1}{N} \sum_{j=1}^{N} 1[X_j^n \leqslant x]$$

where $1[A]$ is the indicator function for the event A. This should approximate the actual cumulative distribution function $F(x)$.

> **Question 1** Write a program to generate such a random sample. Plot the empirical distribution function for $p = 2/3$, using $n = 30$ and N suitably large.

Calculating F

It turns out that for some values of x, we can calculate $F(x)$ explicitly.

> **Question 2** Suppose that
>
> $$x = \sum_{i=1}^{n} \frac{x_i}{2^i}$$
>
> for some $n \in \mathbb{N}$ and some sequence $x_1, \dots, x_n$. (When this is so, we say x has a finite binary expansion.) Find a formula for $F(x)$.

> **Question 3** Use your formula to plot a graph of F, for $p = 3/4$ and $n = 11$, sampling $F(x)$ at $x = 0, 1/2^n, 2/2^n, 3/2^n, 4/2^n, \dots, 1$. Comment briefly on how this graph compares to the graph you obtained in Question 1. Include also a short comparison of the complexity (number of time steps needed) of both algorithms for general n and N.

Properties of F

The plots you have produced should make you wonder: is F continuous? Is it differentiable?

> **Question 4** Let c have a finite binary expansion. Prove that $F(x)$ is continuous at $x = c$. Do your plots suggest that F is continuous elsewhere? Prove or disprove.

We say that F is left-differentiable at c if the limit

$$\lim_{\delta \uparrow 0} \frac{F(c+\delta) - F(c)}{\delta}$$

exists and is finite, and that it is right-differentiable if the limit

$$\lim_{\delta \downarrow 0} \frac{F(c+\delta) - F(c)}{\delta}$$

exists and is finite. If F is both left-differentiable and right-differentiable at c and the two limits are equal, we say F is differentiable at c.

> **Question 5** Let $p = 3/4$ and $c = 9/16$. Plot $(F(c+\delta) - F(c))/\delta$ against δ for a suitable range of values of δ for which $c + \delta$ has a finite binary expansion. Does your plot suggest that F is left-differentiable or right-differentiable at c?

> **Question 6** Make a conjecture about whether F is left-differentiable and/or right-differentiable at an arbitrary point c with a finite binary expansion, for arbitrary $p \in (0, 1)$. Generate two or three plots which support your conjecture. Prove your conjecture.

Project 1.1: Random Binary Expansions

Marking Scheme and additional comments for the Project Report

The purpose of these additional comments is to provide guidance on the structure and length of your CATAM report. Use the same concepts to write the rest of the reports. To help you assess where marks have been lost, this marking scheme will be completed and returned to you during Lent Term. You are advised to keep a copy of your write-up in order to correlate your answers to the marks awarded.

Question no.	**marks available**[1]	**marks awarded**[2]
Programming task *Program:* for instructions regarding printouts and what needs to be in the write-up, refer to the introduction to the manual.		
Question 1 *Comments:* Explain why you are satisfied that your choice of N is reasonable [*approx. 5 lines*][3]	C2.5	
Question 2 *Comments:* Do not include trivial steps in your answer [approx 10 lines] [3]	M1	
Question 3 *Comments:* Plot the graph obtained from the exact values of F, and use a separate graph (or two) for the comparison to the empirical distribution. In your comments try to argue quantitatively if possible. [*approx. 5 lines*][3]. Argue briefly for the complexity [*approx. 3 lines*][3]	C1.5+ M1	
Question 4 *Comments:* Do not include trivial steps in your answer [*approx.1 page*][3]	M3	
Question 5 *Comments:* Scale the axes appropriately, and include a brief justification of your choice. [*approx. 2 lines*][3]	C2	
Question 6 *Comments:* Do not include trivial steps in your proof. [*approx. 1 page*][3]	C2+M5	
Excellence marks. These are awarded for, *among other things*, mathematical clarity and good, clear output (graphs and tables) — see the introduction to the Project Manual.	E2	
Total Raw Marks	20	
Total Tripos Marks	**40**	

[1] C#, M# and E#: *Computational*, *Mathematical* and *Excellence* marks respectively.
[2] For use by the assessor
[3] This figure is only meant to be indicative of the length of your answer, rather than the exact number of lines you are expected to write

