4 Cryptography (mgk25)

In the following, the *index of coincidence* is the probability that two letters picked uniformly at random from different positions in a sequence of letters are identical.

(a) You intercept the following 20-letter sequence:

`CREATETHISNEATDEBATE`

Calculate the index of coincidence for this sequence. [3 marks]

(b) Calculate the expected value of the index of coincidence for letter sequences where the probability of a letter appearing at any given position is

$$p_{\mathrm{E}} = 5/20,\ p_{\mathrm{T}} = 4/20,\ p_{\mathrm{A}} = 3/20,$$
$$\text{and } p_a = 1/20 \text{ for eight other letters } a \in \{\mathtt{B}, \mathtt{C}, \mathtt{D}, \mathtt{H}, \mathtt{I}, \mathtt{N}, \mathtt{R}, \mathtt{S}\}.$$

[*Hint:* the answer will be different from the one for Part (a).] [3 marks]

(c) What is the expected value of the index of coincidence for a letter sequence where each letter was chosen independently uniformly at random from the alphabet $\{\mathtt{A}, \ldots, \mathtt{Z}\}$? [2 marks]

(d) How is the index of coincidence of a letter sequence affected if it is encrypted with a

(i) shift cipher, [2 marks]

(ii) transposition cipher, [2 marks]

(iii) monoalphabetic substitution cipher, [2 marks]

(iv) Vigenère cipher? [2 marks]

Briefly justify your answers.

(e) How can the index of coincidence help to identify a likely key length used in the Vigenère cipher? [4 marks]