EGT3
ENGINEERING TRIPOS PART IIB

---

Monday 6 May 2019 9.30 to 11.10

---

**Module 4F5**

**ADVANCED INFORMATION THEORY AND CODING**

*Answer not more than **three** questions.*

*All questions carry the same number of marks.*

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

**STATIONERY REQUIREMENTS**
Single-sided script paper

**SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM**
CUED approved calculator allowed
Engineering Data Book

**10 minutes reading time is allowed for this paper at the start of the exam.**

**You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.**

1 (a) What are the two least significant decimal digits of $2^{100}$? [10%]

(b) Compute Euler's function $\varphi(56)$. [10%]

(c) (i) Show that if $x$ and $y$ are co-prime integers, any integer $n$ can be written as

$$n = ax + by$$

where $a$ and $b$ are integers. [10%]

(ii) Let $x = 6$ and $y = 9$, which are not co-prime and hence don't satisfy the conditions in part (c)(i). Show that $n = 11$ cannot be expressed as $n = ax + by$ for any integers $a$ and $b$. [10%]

(d) Consider the multiplicative monoid $\langle \mathbb{Z}^{\star}_{105}, \odot \rangle$ of integers $\{1, 2, \ldots, 104\}$ with multiplication modulo 105.

(i) Give an example of an element of $\mathbb{Z}^{\star}_{105}$ that has no multiplicative inverse. [10%]

(ii) Give the inverse of 44, i.e., an element $x \in \mathbb{Z}^{\star}_{105}$ such that $44x$ modulo 105 equals 1. [10%]

(iii) How many invertible elements are there in $\langle \mathbb{Z}^{\star}_{105}, \odot \rangle$? [10%]

(e) Consider the Galois field GF(128) using polynomial arithmetic modulo the primitive polynomial $\pi(X) = 1 + X^3 + X^7$.

(i) What is the multiplicative order of the element $1 + X$ in the field? [10%]

(ii) Calculate $X^7, X^{14}, X^{15}, X^{30}, X^{31}, X^{62}$ and determine the multiplicative inverse of $1 + X$. [10%]

(iii) A linear code is defined over GF(128) by specifying a 5 by 11 parity-check matrix of full row rank. What is the code length and dimension of the equivalent binary code? [10%]

2 A Reed-Solomon code is to be operated over the Galois field GF(101), where $\alpha = 2$ generates this Galois field.

(a) What are the possible code lengths for a Reed-Solomon code over GF(101)? [10%]

(b) If the definition of the Reed-Solomon parity-check matrix is based on the element $\beta = 4$ of GF(101), what would be the resulting code length? [10%]

(c) Now $\beta = 10$ is chosen. Write out the Discrete Fourier Transform (DFT) matrix and its inverse. Note that $1/N = 76$ in GF(101), where $N$ is the DFT length. [15%]

(d) Specify a parity-check matrix for the Reed-Solomon code of rate $R = 1/2$. [5%]

(e) How many codewords does the Reed-Solomon code have? [10%]

(f) How many errors can the Reed-Solomon code detect? [5%]

(g) How many errors can the Reed-Solomon code correct? [5%]

(h) A codeword was obtained by appending two information symbols to a vector of zeros, and taking its inverse DFT. Specify the encoding matrix corresponding to this operation. [10%]

(i) The codeword in part (h) is transmitted through a noisy channel that made a number of transmission errors smaller or equal to the maximum number of errors determined in part (g), and the sequence [91, 10, 73, 30] was received by the decoder. What were the two information symbols encoded? [20%]

(j) The code will be used on a discrete memoryless channel with an error probability of $p = 0.1$ per transmitted symbol of GF(101). What is the probability of successfully decoding a codeword, assuming that decoding will always fail if the number of errors exceeds the maximum computed in part (g)? [10%]

3 (a) The Rivest-Shamir-Adelmann (RSA) cryptosystem is a public key scheme that operates by publishing a pair of integers $(m, e)$ that can be used to encode messages only decryptable by a party who knows the factorisation of $m$ into two large primes $p_1$ and $p_2$. Here, the operation of RSA using *small* primes $p_1 = 11$ and $p_2 = 17$ is considered.

(i) Encrypt the secret message $X = 22$ using the public key $(m, 13)$. [15%]

(ii) Decrypt the public message using the secret key.
It may be helpful for you to note that $\gcd(160, 13) = 1 = 37 \times 13 - 3 \times 160$. [15%]

(b) The Diffie-Hellman key agreement protocol is operated on the multiplicative group $\langle \mathbb{Z}_{101}, \odot \rangle$ using the primitive root $g = 2$.

(i) Alice picks the secret key $a = 16$. Compute her public key $A$. Explain why it is computationally difficult for anyone who knows Alice's public key to discover Alice's secret key. [10%]

(ii) Bob publishes the public key $B = 11$. Compute Alice and Bob's shared secret key. Explain how Bob can compute the same key even though he doesn't know Alice's secret key. [10%]

(c) Consider independent and identically distributed (IID) data $x_1^n = x_1, x_2, \dots, x_n$. In order to test whether they have distribution $P_1$ or $P_2$, a simple likelihood ratio test is employed:

If $P_1^n(x_1^n) > P_2^n(x_1^n)$, declare their distribution is $P_1$; otherwise, declare it is $P_2$.

Now suppose that the data actually comes from a third distribution $Q$, where all three distributions $P_1, P_2$ and $Q$ are different. [So that the test will always give an incorrect answer.]

(i) Assuming $X_1^n \sim Q^n$, find the limit, as $n \to \infty$, of the normalized log-likelihood ratio:

$$\frac{1}{n} \log \frac{P_1^n(X_1^n)}{P_2^n(X_1^n)}.$$

[25%]

(ii) Give conditions on $P_1, P_2$ and $Q$, in terms of relative entropy, characterizing when the result of the test will eventually be $P_1$ or $P_2$, with probability close to 1. [You can ignore the borderline case when the likelihood ratio will be near 1 with high probability.] [25%]

4 Consider the parametric family $\mathcal{P} = \{P_\theta \sim \text{Geom}(\theta) \ : \ \theta \in (0,1)\}$, where the Geom($\theta$) distribution has probability mass function $P_\theta(k) = \theta(1-\theta)^{k-1}$, for $k = 1, 2, \ldots$, with mean $1/\theta$ and variance $(1-\theta)/\theta^2$. Suppose the samples $x_1^n$ are generated by the independent and identically distributed $X_1^n$, distributed according to some $P_\theta \in \mathcal{P}$.

(a) Show that the maximum likelihood estimate (MLE) $\hat{\theta}_{\text{MLE}}(x_1^n)$ for $\theta$ is:

$$\hat{\theta}_{\text{MLE}}(x_1^n) = \left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]^{-1}.$$

[20%]

(b) Prove that the MLE is biased:

$$E_\theta[\hat{\theta}_{\text{MLE}}(X_1^n)] > \theta, \quad \text{for all } \theta \in (0,1).$$

[20%]

(c) Compute the Fisher information $J(\theta)$ for the family $\mathcal{P}$. [20%]

(d) Compute the bias of $\hat{\theta}_{\text{MLE}}$ in the case of a single sample, $n = 1$.
It may be helpful to note that the Taylor series expansion for the natural logarithm is:

$$\log_e(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}, \quad x \in (0,1).$$

[20%]

(e) Find a lower bound for the mean-squared error $\text{MSE}(\hat{\theta}_{\text{MLE}}; \theta)$ achieved by $\hat{\theta}_{\text{MLE}}$ in the case of a single sample. Express the bound as a function of $\theta$; you do not need to simplify the resulting expression. [20%]

**END OF PAPER**

THIS PAGE IS BLANK