

**Digital Circuits**  
**Midterm, Tuesday May 2nd**

**Instructions:**

- You have **two hours**, 7PM - 9PM
- The exam has 3 questions, totaling 100 points.
- Please start answering each question on a new page of the answer booklet.
- You are allowed to carry the textbook, your own notes and other course related material with you. Electronic reading devices [including kindles, laptops, ipads, etc.] are allowed, provided they are used solely for reading pdf files already stored on them and not for any other form of communication or information retrieval.
- You are required to provide a detailed explanation of how you arrived at your answers.
- You can use previous parts of a problem even if you did not solve them.
- As throughout the course, entropy ( $H$ ) and Mutual Information ( $I$ ) are specified in bits.
- $\log$  is taken in base 2.
- Throughout the exam 'prefix code' refers to a variable length code satisfying the prefix condition.
- Good Luck!

# 1. Mix of Questions (40 points)

You only need to answer **four out of the five** questions presented below. Each of them is worth 10 points.

- 1) Let  $Z_1, Z_2, Z_3, \dots$  be i.i.d. random variables that take values “0” and “1” with equal probability. Further, let

$$X_i = \sum_{j=1}^i Z_j, \text{ for } 1 \leq i \leq n. \quad (1)$$

Find  $I(X_1; X_2, X_3, \dots, X_n)$ .

- 2) Let  $U_1, U_2, U_3, \dots$  be i.i.d. taking values  $A, B, C, D, E$  and  $F$ , with the following distribution:

Symbol	$A$	$B$	$C$	$D$	$E$	$F$
Probability	$1/2$	$1/4$	$1/8$	$1/16$	$1/32$	$1/32$

- (a) Compute  $H(U_1)$ .  
 (b) What is the most probable sequence of a given length  $n$ ? What is its probability?  
 (c) Recall the definition of the  $\epsilon$ -typical set for a memoryless source  $U$ :

$$A_\epsilon^{(n)} = \left\{ u^n : \left| -\frac{1}{n} \log p(u^n) - H(U) \right| \leq \epsilon \right\}. \quad (2)$$

Does the sequence you found in part (b) belong to  $A_\epsilon^{(n)}$  for  $\epsilon = 0.1$ ? How about for  $\epsilon = 1$ ?

- 3) Let  $(X_i, Y_i)$  be i.i.d.  $\sim p(x, y)$ . Find the limit in probability, as  $n \rightarrow \infty$ , of

$$\frac{1}{n} \log \frac{p(X^n, Y^n)}{p(X^n)p(Y^n)}. \quad (3)$$

- 4) Consider a source with five symbols  $u_1, u_2, u_3, u_4, u_5$ , with probabilities  $p(u_1) \geq p(u_2) \geq p(u_3) \geq p(u_4) \geq p(u_5)$ .

- (a) Suppose  $p(u_1) \geq p(u_2) = p(u_3) = p(u_4) = p(u_5)$ . Find the minimum value of  $q$  such that  $p(u_1) \geq q$  implies  $n_1 = 1$ . Here  $n_1$  denotes the length of the codeword associated with symbol  $u_1$  generated by a Huffman code applied to the source.  
 (b) Suppose  $p(u_1) \geq p(u_2) \geq p(u_3) \geq p(u_4) > p(u_5) = 0$ . Find the largest value of  $r$  such that  $p(u_1) \leq r$  implies  $n_1 > 1$ . Here  $n_1$  denotes the length of the codeword associated with symbol  $u_1$  generated by a Huffman code applied to the source.

- 5) Consider a random variable  $X$  which takes on four possible values with probabilities  $(1/3, 1/3, 1/4, 1/12)$ .
- (a) Construct a Huffman code for this random variable.
  - (b) Show that there exist two different sets of optimal lengths for the codewords, namely, show that codeword length assignments  $(1, 2, 3, 3)$  and  $(2, 2, 2, 2)$  are both optimal.
  - (c) Are there optimal codes with codeword lengths for some symbols that exceed the Shannon code length  $\lceil \log \frac{1}{p(x)} \rceil$ ? (Hint: Check the codeword lengths from the previous part.)

## 2. Non-prefix Code (30 points)

Suppose  $p(x)$  is a PMF over  $\mathcal{X} = \{1, 2, \dots, K\}$ , with  $p(1) > p(2) > \dots > p(K)$ . We want to encode a random variable  $X \sim p$ . We care about encoding only this one random variable, therefore we do not require Unique Decodability but merely that the code be one-to-one, i.e., a different codeword for each of the  $K$  source symbols. Note that even the zero length codeword is valid, i.e., sending nothing (the empty string) can represent one of the source symbols.

- (a) (5 points) Construct a coding scheme  $c(X)$  that has the minimum expected code length. Let  $l(i)$  be the length of the codeword of symbol  $i$ . Show that  $l(i) = \lfloor \log i \rfloor$ , where  $\lfloor a \rfloor$  is the greatest integer no bigger than  $a$ .
- (b) (10 points) Prove that the coding scheme from Part (a) satisfies

$$l(i) \leq -\log p(i)$$

and conclude that the minimum expected code length is less than or equal to the entropy, i.e.,

$$\mathbb{E}[l(X)] \leq H(X).$$

[Hint : Note that  $p(i)$  is the  $i$ -th largest value, and therefore,  $P(i) \leq \frac{1}{i}$ ]

- (c) (15 points) Show that

$$\mathbb{E}[l(X)] \geq H(X) - 1 - \log(1 + \log K).$$

That is, lossless codes, even if not Uniquely Decodable, cannot beat the entropy by much.

[Hint : You may want to use the fact that  $\sum_{i=1}^K \frac{1}{i} \leq 1 + \log K$ ]

### 3. The prime number theorem (30 points)

Some time around 300 B.C., someone showed that there are infinitely many prime numbers – we know this because a proof appears in Euclid’s famous *Elements*. In this problem, we will not only show that there are infinitely many prime numbers, but we will also give a lower bound on the rate of their growth using information theory.

Let  $\pi(n)$  denote the number of primes no greater than  $n$ . Note that every positive integer  $n$  has a **unique** prime factorization of the form

$$n = \prod_{i=1}^{\pi(n)} p_i^{X_i}, \quad (4)$$

where  $p_1, p_2, p_3, \dots$  are the primes, that is,  $p_1 = 2$ ,  $p_2 = 3$ ,  $p_3 = 5$ , etc., and  $X_i = X_i(n)$  is the non-negative integer representing the multiplicity of  $p_i$  in the prime factorization of  $n$ .

Let  $N$  be uniformly distributed on  $\{1, 2, 3 \dots n\}$ .

(a) (8 points) Show that  $X_i(N)$  is an integer-valued random variable satisfying

$$0 \leq X_i(N) \leq \log n. \quad (5)$$

[Hint : Try finding a lower and an upper bound for  $p_i^{X_i(N)}$ ]

(b) (22 points) Show that

$$\log n = H(N) \leq \pi(n) \log(\log n + 1). \quad (6)$$

Thus, not only is  $\pi(n) \rightarrow \infty$  but in fact  $\pi(n) \geq \frac{\log n}{\log(\log n + 1)}$ .

[Hint : Do  $X_1(N), X_2(N), \dots, X_{\pi(n)}(N)$  determine  $N$ ? What does that say about the respective entropies?].