

# REPORT

February 4, 2025

---

Practice Exercises for Information-Theory 2

---

Professor Petros Elia, elia@eurecom.fr

February 9th, 2024

---

**1 Let  $\mathbb{F}_n^2$  be the set of all binary vectors, and let  $\mathcal{A}_\epsilon^n$  be the set of typical sequences based on non-trivial distribution. Which is true below?**

- a)  $Prob(\mathcal{A}_\epsilon^n) \approx 1$  or b)  $Prob(\mathcal{A}_\epsilon^n) \ll 1$
- c)  $|\mathcal{A}_\epsilon^n| \ll 2^n$  or d)  $|\mathcal{A}_\epsilon^n| \approx 1$  ?

In the above, choose a) or b) and then choose c) or d). Justify.

---

For this problem, we are dealing with the typical set  $\mathcal{A}_\epsilon^n$  from information theory, specifically from the context of the **Asymptotic Equipartition Property (AEP)**.

**Part 1: Probability of the typical set** The typical set  $\mathcal{A}_\epsilon^n$  is defined such that most of the probability mass is concentrated there as  $n$  becomes large. Based on the AEP, for large  $n$ , the probability of drawing a sequence from the typical set approaches 1:

$$P(\mathcal{A}_\epsilon^n) \approx 1$$

This corresponds to option **a)**. Therefore, **a)** is correct.

**Part 2: Size of the typical set** Now, let's consider the size of the typical set. For a random variable with entropy  $H(X)$ , the number of typical sequences is roughly  $2^{nH(X)}$ . Since  $H(X) < \log_2(2^n) = n$  (entropy is always less than or equal to  $n$  for a binary source), we conclude that the size of the typical set is much smaller than the total number of sequences in  $\mathbb{F}_n^2$ , which is  $2^n$ .

Therefore:

$$|\mathcal{A}_\epsilon^n| \ll 2^n$$

This corresponds to option **c)**.

**Conclusion:** The correct answers are: - **a)**:  $P(\mathcal{A}_\epsilon^n) \approx 1$  - **c)**:  $|\mathcal{A}_\epsilon^n| \ll 2^n$

These are consistent with the properties of the typical set in information theory.

- 2 In a communication setting where  $X$  defines the input and  $Y$  defines the output, what is the connection between the mutual information  $I(X;Y)$  and the channel capacity? Offer some intuition as well as review the proof.
- 

### Connection Between Mutual Information $I(X;Y)$ and Channel Capacity $C$

1. **Definitions:**

- **Mutual Information:**  $I(X;Y) = H(X) - H(X|Y)$ 
  - Measures the amount of information that  $Y$  reveals about  $X$ .
- **Channel Capacity:**  $C = \max_{p(x)} I(X;Y)$ 
  - The maximum rate at which information can be reliably transmitted over a channel.

2. **Intuition:**

- $I(X;Y)$  quantifies the reduction in uncertainty about  $X$  when  $Y$  is known.
- $C$  is achieved by selecting the input distribution  $p(x)$  that maximizes  $I(X;Y)$ .

3. **Key Steps in the Proof:**

- **Channel Model:** For a discrete memoryless channel, characterized by transition probabilities  $P(y|x)$ .
- **Maximization:** Find  $C$  by maximizing  $I(X;Y)$  over all input distributions.
- **Achievability:** Shannon's theorem shows that  $C$  can be approached with specific coding strategies, ensuring low error rates.

4. **Conclusion:**

- The relationship between  $I(X;Y)$  and  $C$  is crucial for designing efficient communication systems, where maximizing mutual information leads to reliable transmission rates.

- 3 Consider a multiple-input multiple-output (MIMO) channel with 2 transmit antennas, 3 receive antennas, and random fading where you can only encode over 1 coherence period. Assume Rayleigh fading where the fading coefficients are iid and distributed as complex-normal random variables with zero mean and unit variance. What is approximately the probability of error if the signal-to-noise ratio is approximately SNR= 10000?
- 

Let's break down and analyze the problem step-by-step to find the approximate probability of error.

#### Step 1: Channel Setup and Assumptions

- **MIMO Setup:** 2 transmit antennas, 3 receive antennas.
- **Rayleigh Fading:** Each channel coefficient  $h_{ij}$  is distributed as  $\mathcal{CN}(0,1)$  (complex Gaussian with zero mean and unit variance).
- **Signal-to-noise ratio (SNR):** SNR = 10,000 (i.e., 40 dB).

## Step 2: Error Analysis

- For high SNR scenarios, the probability of error is dominated by the diversity order of the MIMO system.
- **Diversity Order:**  
In MIMO systems, the diversity order is given by the product of the number of transmit and receive antennas,  $D = N_t \times N_r$ . Here:  $D = 2 \times 3 = 6$
- The probability of error in MIMO Rayleigh fading channels can be approximated at high SNR using the formula:  $P_e \approx \left(\frac{1}{\text{SNR}}\right)^D$

**Step 3: Compute the Probability of Error** Given  $\text{SNR} = 10,000 = 10^4$  and  $D = 6$ :  $P_e \approx \left(\frac{1}{10^4}\right)^6 = 10^{-24}$

## Step 4: Interpretation

- The error probability is exceedingly small ( $10^{-24}$ ), which is expected for a system with high diversity and very high SNR.

This result demonstrates how increasing both SNR and diversity significantly reduces the error probability in MIMO systems under Rayleigh fading.

## 4 Can you provide real-life settings (emphasis on the channel behavior) where we must use the metric of ergodic capacity, of $\epsilon$ -outage capacity, and of AWGN capacity?

---

### 4.0.1 Capacity Metrics and Channel Settings

Metric	Channel Behavior	Real-Life Example
<b>Ergodic Capacity</b>	Rapid, time-varying fading	Mobile broadband (e.g., 4G/5G urban)
<b><math>\epsilon</math>-Outage Capacity</b>	Deep fades, unreliable links	Vehicle-to-Everything (V2X) safety
<b>AWGN Capacity</b>	Stable, no fading	Satellite communication

- **Ergodic:** Used when channel variations average out over time.
- **Outage:** Used where reliability is crucial despite deep fades.
- **AWGN:** Used when channel is static and predictable.

## 5 Explain the difference between the orthogonal and the non-orthogonal multiple-access settings. What are the practical consequences on the receiver?

---

### 5.0.1 Orthogonal vs Non-Orthogonal Multiple Access

Feature	Orthogonal Access (OMA)	Non-Orthogonal Access (NOMA)
<b>Resource Allocation</b>	Separate, non-overlapping (e.g., time, freq.)	Shared resources (e.g., same time & freq.)
<b>Interference</b>	No user-to-user interference	Interference between users
<b>Receiver Complexity</b>	Low (no interference cancellation)	High (requires SIC or similar techniques)
<b>Resource Efficiency</b>	Lower (may have idle resources)	Higher (better resource utilization)
<b>Example</b>	OFDMA (e.g., 4G LTE)	NOMA (e.g., 5G)

- **OMA:** Simple receivers but lower efficiency.
- **NOMA:** Efficient but complex receivers due to interference management.

**6 Consider a multi-antenna communication setting. Let us be able to send 10 different 64-QAM symbols in 4 channel uses. What is the rate of communication in bits per channel use (bpcu)? What would the constellation size be if, under the same conditions, we wished to double the rate.**

#### 6.0.1 Step 1: Calculate the rate in bits per channel use (bpcu)

Given: - **Symbols:** 10 symbols transmitted  
- **Constellation:** 64-QAM (each symbol represents  $\log_2(64) = 6$  bits)  
- **Channel uses:** 4

**Rate Calculation** The total number of bits sent: Bits sent =  $10 \times 6 = 60$  bits

The rate in bits per channel use (bpcu) is:  $R = \frac{\text{Bits sent}}{\text{Channel uses}} = \frac{60}{4} = 15$  bpcu

#### 6.0.2 Step 2: Doubling the rate

We want to double the rate to:  $R' = 2 \times 15 = 30$  bpcu

#### 6.0.3 Step 3: Find the new constellation size

Let  $M'$  be the new constellation size. We still send **10 symbols** in **4 channel uses**, so the required bits per symbol are: Bits per symbol =  $\frac{30 \text{ bpcu} \times 4 \text{ channel uses}}{10 \text{ symbols}} = 12$  bits per symbol

The constellation size is:  $M' = 2^{12} = 4096$

---

#### 6.0.4 Final Answer

- **Current rate:** 15 bpcu
- **New constellation size for doubled rate:**  $M' = 4096$ -QAM

**7** Consider gathering statistics on the frequency with which different 6-tuples of letters appear in the English language. Argue what is (the order of, i.e., approximately) the amount of data that you'd need to gather such statistics that are reliable.

---

**Step 1: Number of Possible 6-tuples** For a 6-tuple of letters, assuming there are 26 letters in the English alphabet, the total number of possible combinations is:  $N = 26^6 = 308,915,776$

**Step 2: Statistical Reliability** To gather reliable statistics, each 6-tuple should ideally appear multiple times (e.g., at least **10 times** for reasonable statistical confidence). However, most 6-tuples will occur rarely, as English has strong constraints based on word structure, making many tuples highly improbable.

Let  $f_{\min}$  denote the frequency of the **least common relevant tuple**.

In natural English text: - Common 6-tuples (e.g., “that is”, “should”) occur frequently. - Rare ones (e.g., “qzxvwy”) may rarely appear, if ever.

To approximate data needs: - Suppose  $N_{\text{effective}}$ , the number of realistically occurring 6-tuples, is approximately  $10^6$  based on linguistic constraints. - We want **at least 10 occurrences** of each relevant tuple: Required data size  $\approx 10 \times N_{\text{effective}} = 10^7$  tuples

**Step 3: Data in English Text** Assuming about **5 characters per word** on average (including spaces/punctuation): Words required  $\approx 2 \times 10^6$  words

In terms of text: - Average English text has ~250 words per page. - Data requirement =  $\frac{2 \times 10^6 \text{ words}}{250 \text{ words per page}} \approx 8000$  pages.

#### Summary

- **Order of data required:**  $\sim 10^7$  6-tuples or  $\sim 2$  million words of English text ( $\sim 8000$  pages).
- This ensures each 6-tuple appears enough times to gather reliable statistics.

- 8 Consider the 3-user broadcast channel, with links  $|h_1|^2 = 0.7, |h_2|^2 = 0.4, |h_3|^2 = 0.2$ , having also a unit power noise at the receivers. Let  $P_k$  be the power of the signal meant for user  $k$ , and let  $P_1 = 10, P_2 = 5$  and  $P_3 = 1$ . Describe the capacity region, and how this is achieved by describing the method of decoding.

#### 8.0.1 Capacity Region of 3-User Broadcast Channel

User	Channel Gain $ h_k ^2$	Power $P_k$	SNR $\frac{ h_k ^2 P_k}{1}$	Achievable Rate $R_k$
User 1	0.7	10	7	$\log_2(1 + 7) \approx 3$ bpcu
User 2	0.4	5	2	$\log_2(1 + 2) \approx 1.585$ bpcu
User 3	0.2	1	0.2	$\log_2(1 + 0.2) \approx 0.263$ bpcu

#### 8.0.2 Method of Achieving Capacity

1. **Superposition Coding:** The transmitter sends a superposition of all user signals.
2. **Successive Interference Cancellation (SIC):**
  - **User 1:** Decodes  $X_3$ , then  $X_2$ , and finally  $X_1$ .
  - **User 2:** Decodes  $X_3$ , removes it, and then decodes  $X_2$ .
  - **User 3:** Directly decodes  $X_3$  with no interference removal.

#### 8.0.3 Summary

- **Achievable rates:**  $R_1 \approx 3, R_2 \approx 1.585, R_3 \approx 0.263$  bpcu.
- The capacity region is achieved through superposition coding and SIC.

- 9 What is approximately the doubling rate in a horse race, if your starting bet is 10 euros, you always bet your accumulated funds, and your income in 10 bets is equal to 20000 euros?

#### Step 1: Define Variables

- Starting bet:  $B_0 = 10$  euros
- Final accumulated income:  $B_{10} = 20,000$  euros
- We bet our entire funds each time, so the capital grows by a factor per bet.

Let the growth factor per bet be  $r$ . The capital after  $n$  bets is:  $B_n = B_0 \cdot r^n$

For  $n = 10$ :  $B_{10} = 10 \cdot r^{10} = 20,000$

**Step 2: Solve for  $r$**   $r^{10} = \frac{20,000}{10} = 2,000$   $r = (2,000)^{1/10} \approx 3.162$

**Step 3: Calculate Doubling Rate** The doubling rate refers to how many bets it takes for the capital to double. We need to solve for  $n$  such that:  $2 = r^n$   $n = \frac{\log(2)}{\log(r)} = \frac{\log(2)}{\log(3.162)}$

Calculating:  $n \approx \frac{0.3010}{0.4997} \approx 0.602$  bets

**Final Answer**

- The approximate doubling rate is **0.6 bets**.

**10 Derive the Kullback-Leibler distance between two distributions: the first is the fair dice distribution, and the other is the distribution where the equiprobable outcomes 1,2,3 have a double probability than equiprobable outcomes 4,5,6.**

---

**Step 1: Define the two distributions**

1. **Fair dice distribution:**

$$P(i) = \frac{1}{6}, \quad i \in \{1, 2, 3, 4, 5, 6\}$$

2. **Skewed distribution:**

Let the outcomes 1, 2, 3 have **double the probability** of 4, 5, 6.

Define the probabilities in the new distribution  $Q$ :

- For  $i = 1, 2, 3$ , let  $Q(i) = 2p$ .
- For  $i = 4, 5, 6$ , let  $Q(i) = p$ .

Since the sum of all probabilities must equal 1:  $3(2p) + 3(p) = 1 \Rightarrow 9p = 1 \Rightarrow p = \frac{1}{9}$

Therefore, the skewed distribution  $Q$  is:  $Q(i) = \begin{cases} \frac{2}{9}, & i \in \{1, 2, 3\} \\ \frac{1}{9}, & i \in \{4, 5, 6\} \end{cases}$

**Step 2: Kullback-Leibler (KL) Divergence** The KL divergence between  $P$  and  $Q$  is defined as:  $D_{\text{KL}}(P||Q) = \sum_{i=1}^6 P(i) \log \left( \frac{P(i)}{Q(i)} \right)$

For the fair dice distribution  $P(i) = \frac{1}{6}$ , we compute the divergence separately for  $i = 1, 2, 3$  and  $i = 4, 5, 6$ .

**Step 3: Compute the KL divergence**

1. **For  $i = 1, 2, 3$ :**  $P(i) = \frac{1}{6}$ ,  $Q(i) = \frac{2}{9}$  Contribution to KL divergence:  $\frac{1}{6} \log \left( \frac{\frac{1}{6}}{\frac{2}{9}} \right) = \frac{1}{6} \log \left( \frac{9}{12} \right) = \frac{1}{6} \log \left( \frac{3}{4} \right)$
2. **For  $i = 4, 5, 6$ :**  $P(i) = \frac{1}{6}$ ,  $Q(i) = \frac{1}{9}$  Contribution to KL divergence:  $\frac{1}{6} \log \left( \frac{\frac{1}{6}}{\frac{1}{9}} \right) = \frac{1}{6} \log \left( \frac{9}{6} \right) = \frac{1}{6} \log \left( \frac{3}{2} \right)$

**Step 4: Total KL divergence** The total KL divergence is the sum of the contributions from both groups of outcomes:  $D_{\text{KL}}(P||Q) = 3 \cdot \frac{1}{6} \log\left(\frac{3}{4}\right) + 3 \cdot \frac{1}{6} \log\left(\frac{3}{2}\right)$

Simplify:  $D_{\text{KL}}(P||Q) = \frac{1}{2} \left( \log\left(\frac{3}{4}\right) + \log\left(\frac{3}{2}\right) \right) = \frac{1}{2} \log\left(\frac{3}{4} \cdot \frac{3}{2}\right) = \frac{1}{2} \log\left(\frac{9}{8}\right)$

**Final Answer** The Kullback-Leibler divergence between the two distributions is:  $D_{\text{KL}}(P||Q) = \frac{1}{2} \log\left(\frac{9}{8}\right)$

**11 Consider the coded caching scenario, where a transmitter serves  $K = 3$  users via a broadcast shared-link (bottleneck-link), and where the transmitter has access to a library of  $N = 6$  files (movies) A,B,C,D,E,F, each of size 1GigaByte (GB). Let each user have a cache of size  $M = 2$  GBs.**

Describe: - The placement phase (what data goes into each user's cache) - The delivery phase (describe the sequence of XORs sent by the transmitter) - What is the total size of all the transmitted XORs together?

---

**Given:**

- **Number of users:**  $K = 3$
- **Library size:**  $N = 6$  files (A, B, C, D, E, F), each 1 GB in size
- **Cache size per user:**  $M = 2$  GB
- **Goal:** Efficient placement and delivery to minimize data sent over the bottleneck-link

We will follow **Maddah-Ali and Niesen's Coded Caching Scheme** to describe the placement and delivery phases.

**1. Placement Phase** The placement phase is done without knowledge of future user requests. Each file is divided into **subfiles** and cached according to the combinatorial scheme.

Let each file (e.g., file A) be divided into  $\binom{K}{t} = 3$  parts, where  $t = \frac{KM}{N} = 1$ .

Thus, each file  $W$  has **3 subfiles**:

$$W = \{W_1, W_2, W_3\}$$

For each user  $i$ , cache the subfiles corresponding to all subsets of  $\{1, 2, 3\}$  that do **not** include  $i$ . Specifically: - **User 1** caches all subfiles  $W_2$  and  $W_3$  for every file  $W$ . - **User 2** caches all subfiles  $W_1$  and  $W_3$  for every file  $W$ . - **User 3** caches all subfiles  $W_1$  and  $W_2$  for every file  $W$ .

Each user stores  $\frac{2}{3}$  of every file, leading to a total cache size of  $M = 2$  GBs.

**2. Delivery Phase** In this phase, each user requests a file. Assume the following requests: - **User 1:** File A - **User 2:** File B - **User 3:** File C

Since each user already has  $\frac{2}{3}$  of every file, the transmitter only needs to deliver the remaining **missing subfiles** in an efficient way. The missing subfiles are: - User 1 needs  $A_1$  - User 2 needs  $B_2$  - User 3 needs  $C_3$



The transmitter exploits coded multicasting by sending **XORs** of missing subfiles that different users can decode using their cached data: 1. **Send:**  $A_1 \oplus B_2 \oplus C_3$

Each user can decode their missing subfile: - **User 1:** Knows  $B_2$  and  $C_3$ , can extract  $A_1$ . - **User 2:** Knows  $A_1$  and  $C_3$ , can extract  $B_2$ . - **User 3:** Knows  $A_1$  and  $B_2$ , can extract  $C_3$ .

**3. Total Size of Transmitted XORs** Since each XOR combines **1/3 of a file** for each user, the total size of all transmitted XORs is: Total transmitted size =  $\frac{1}{3}$  GB

Since only one XOR transmission is needed in this case, the total size is **1 GB**.

## Summary

### 1. Placement Phase:

- Each user caches  $\frac{2}{3}$  of every file, totaling 2 GB of data.

### 2. Delivery Phase:

- The transmitter sends  $A_1 \oplus B_2 \oplus C_3$ , which each user decodes using cached subfiles.

### 3. Total transmitted size: 1 GB.

**12 Consider the setting of horse-races with  $m=4$  horses, where the odds are even. Let the probability that the first horse wins be  $p_1 = 1/2$ , and the rest equal to  $1/6$ . Assume you bet 100 euros (starting bet).**

- To be optimal, how much should you bet on the fourth horse (horse number 4)?
- What is the optimal doubling rate?

---

**Given:**

- **Number of horses:**  $m = 4$
- **Odds:** Even (1-to-1 payout, win doubles the bet)
- **Probabilities:**
  - $p_1 = \frac{1}{2}$ ,
  - $p_2 = p_3 = p_4 = \frac{1}{6}$
- **Starting bet:** 100 euros
- **Goal:** Maximize the doubling rate  $\mathcal{D}$ .

**1. Optimal Betting Strategy** We apply **Kelly's Criterion**, which maximizes the long-term growth rate of capital. Let  $f_k$  be the fraction of the capital bet on horse  $k$ , with the constraint:  $\sum_{k=1}^4 f_k = 1$

The optimal bets are determined by maximizing the **expected log wealth**:  $\mathcal{D} = \sum_{k=1}^4 p_k \log(1 + 2f_k)$

**Step 1: Optimal Bet Fractions** The optimal fraction  $f_k$  for horse  $k$  is:  $f_k = \frac{p_k}{1+p_k}$

1. **For horse 1:**

$$p_1 = \frac{1}{2} \quad f_1 = \frac{\frac{1}{2}}{1+\frac{1}{2}} = \frac{1}{3}$$

2. **For horse 2, 3, and 4:**

$$p_2 = p_3 = p_4 = \frac{1}{6} \quad f_2 = f_3 = f_4 = \frac{\frac{1}{6}}{1+\frac{1}{6}} = \frac{1}{7}$$

**Step 2: Amount to Bet on Horse 4**

- **Fraction to bet on horse 4:**  $f_4 = \frac{1}{7}$
- **Total capital:** 100 euros
- **Bet on horse 4:**  
Bet amount =  $f_4 \times 100 = \frac{1}{7} \times 100 \approx 14.29$  euros

**Step 3: Optimal Doubling Rate** The optimal doubling rate  $\mathcal{D}$  is:  $\mathcal{D} = \sum_{k=1}^4 p_k \log(1 + 2f_k)$

1. **For horse 1:**

$$\mathcal{D}_1 = \frac{1}{2} \log\left(1 + 2 \times \frac{1}{3}\right) = \frac{1}{2} \log\left(\frac{5}{3}\right)$$

2. **For horses 2, 3, and 4:**

$$\mathcal{D}_2 = \frac{1}{6} \log\left(1 + 2 \times \frac{1}{7}\right) = \frac{1}{6} \log\left(\frac{9}{7}\right)$$

Thus, the total doubling rate is:  $\mathcal{D} = \frac{1}{2} \log\left(\frac{5}{3}\right) + 3 \cdot \frac{1}{6} \log\left(\frac{9}{7}\right)$

**Final Answers:**

- **Bet on horse 4: 14.29 euros**
- **Optimal doubling rate:**  
 $\mathcal{D} = \frac{1}{2} \log\left(\frac{5}{3}\right) + \frac{1}{2} \log\left(\frac{9}{7}\right) \approx 0.238$  bpcu

**13 What is the entropy of a zero-mean Real Gaussian Random Variable with variance 1. How does the entropy change if the variance doubles, and how if the mean shifts to 3?**

---

**Entropy of a Real Gaussian Random Variable** For a **zero-mean** real Gaussian random variable  $X$  with variance  $\sigma^2$ , the entropy is given by:  $H(X) = \frac{1}{2} \log(2\pi e \sigma^2)$

**Step 1: Entropy for  $\sigma^2 = 1$**  When  $X \sim \mathcal{N}(0,1)$ ,  $\sigma^2 = 1$ , the entropy is:  $H(X) = \frac{1}{2} \log(2\pi e \cdot 1) = \frac{1}{2} \log(2\pi e)$

**Step 2: If the variance doubles** If the variance becomes  $\sigma^2 = 2$ , the entropy is:  $H(X) = \frac{1}{2} \log(2\pi e \cdot 2) = \frac{1}{2} \log(4\pi e)$

Change in entropy:  $\Delta H = \frac{1}{2} \log\left(\frac{4\pi e}{2\pi e}\right) = \frac{1}{2} \log(2)$

The entropy increases by  $\frac{1}{2} \log(2) \approx 0.347$  **bits**.

**Step 3: If the mean shifts to 3** Entropy is invariant under mean shifts. If the mean shifts from 0 to 3, the entropy remains:  $H(X) = \frac{1}{2} \log(2\pi e \cdot \sigma^2)$

**Change in entropy: 0 bits.**

#### Summary

- **Entropy of  $\mathcal{N}(0, 1)$ :**  $\frac{1}{2} \log(2\pi e) \approx 1.418$  bits
- **If the variance doubles:** Entropy increases by  $\frac{1}{2} \log(2) \approx 0.347$  bits
- **If the mean shifts to 3:** Entropy does not change.

[ ]: