# Problem Set

Lakshmi Natarajan*

*Abstract*—**Problems discussed in the EE 2340 course on Information Theory, taught at IIT Hyderabad, are listed in this document. The problems cover all the basic concepts in Information Theory.**

1) Suppose $X$ and $Y$ have the following joint probability mass function $P_{XY}(0,0) = P_{XY}(0,1) = P_{XY}(1,1) = \frac{1}{3}$ and $P_{XY}(1,0) = 0$ [1]
   (a) Find the marginal distributions $P(X)$ and $P(Y)$.
   (b) Determine if $X$ and $Y$ are independent.
   (c) Find the conditional distributions $P(X|Y)$ and $P(Y|X)$.
   (d) Find $H(X)$,$H(Y)$ and $H(X,Y)$.
   (e) Find $H(X|Y)$,$H(Y|X)$ and $I(X;Y)$.
2) Show that, for any three random variables $X$,$Y$,$Z$,
$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z).$$
3) Let $X$ and $Y$ be real-valued discrete random variables with joint probability mass function $p(x,y)$, and let $Z = X + Y$. Show that $H(Z|X) = H(Y|X)$.
4) Let the vector $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ denote the probability mass function of a discrete random variable $X$. In particular, $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. Show that $H(X) = 0$ if and only if exactly one of $p_1, ..., p_n$ is non-zero.
   *Remark:* This implies there exists an $i$ such that $p_i = 1$ and $p_j = 0$ for all $j \neq i$.
5) Let $X$ be a random variable with alphabet $\mathcal{X} = \{x_1, ..., x_n\}$ and $g : \mathcal{X} \to \mathcal{Y}$ be any function, where $\mathcal{Y}$ is any set. Note that $g(X)$ takes values $g(x_1), ..., g(x_n) \in \mathcal{Y}$ [1]
   (a) Find $P(g(X) = g(x_i)|X = x_j)$ for $i, j \in \{1, ..., n\}$.
   (b) Show that $H(g(X)|X) = 0$.
   (c) Using the chain rule of entropy on $H(X, g(X))$ show that $H(g(X)) \leq H(X)$.
6) Answer briefly:
   (a) Write the expression for the Kullback-Leibler divergence $D(p \parallel q)$ between two probability distributions $p(x)$ and $q(x)$.
   (b) State the log-sum inequality. When is equality attained in the log-sum inequality?
   (c) Determine if the function $\log(x^3)$ is convex/strictly convex/concave/strictly concave/none of the above in the interval $0 < x < \infty$.
   (d) For any random variable $X$, what is the value of $I(X;X)$?
7) Suppose it is possible to determine the value of a random variable $X$ from $Y$, and the value of $Y$ from $X$, with probability 1. Show that $H(X) = H(Y)$.
8) Let $X$ and $Z$ be real valued random variables and let $X$ be independent of $Z$. If $Y = X + Z$, then show that $H(Y|X) = H(Z)$.
9) Suppose $X$ and $X'$ are independent random variables with probability mass functions $p(x)$ and $q(x)$, respectively, where $x \in \mathcal{X}$ [2]
   (a) In terms of the functions $p(x)$ and $q(x)$, find the expression for the probability that $X = X'$, i.e, $P(X = X')$.
   (b) Show that the cross entropy $\sum_{x \in \mathcal{X}} p(x) \log(\frac{1}{q(x)})$ is equal to the sum $H(X) + D(p \parallel q)$.
   (c) Use Jensen's inequality to show that
$$-\log(P(X = X')) \leq H(X) + D(p \parallel q).$$
   *Remark:* This result provides a lower bound $P(X = X') \geq 2^{-H(X) - D(p \| q)}$ in terms of entropy of $X$ and the $KL$ divergence between the distributions of $X$ and $X'$.

The author is with the Department of Electrical Engineering, IIT Hyderabad, Kandi 502285.

10) Let $X = \{1, ..., n\}$ and $P(X = i) = p_i$ for $i = 1, ..., n$. Suppose $k$ is some fixed integer with $1 < k < n$. Define a random variable $Z$ such that $Z = 0$ if $X \leq k$ and $Z = 1$ if $X > k$. Derive the expression of $H(X|Z)$ in terms of $p_1, ..., p_n$.
11) *Rate-Distortion Theory(compressing a source with distortion) - An example:* Let $X$ be a random binary vector of length $n$, i.e, its alphabet $X = \{0,1\}^n$ is the set of all binary vectors of length $n$. Assume that $X$ is uniformly distributed over $X$. Let $Y = f(X)$ be a length $l$ binary vector obtained by 'encoding' $X$ using a known function $f$, i.e, $Y = \{0,1\}^l$ and $f : X \to Y$. Suppose an estimate $\hat{X} \in X$ of $X$ is obtained by using a 'decoding' function $g : Y \to X$, i.e, $\hat{X} = g(Y) = g(f(X))$. Assume that the vector $\hat{X}$ differs from $X$ in at the most 1 position,i.e, either $\hat{X} = X$ or $\hat{X}$ differs from $X$ in exactly one of the $n$ coordinates. Derive the best possible bounds in the following questions.
    (a) Derive an upper bound on $H(X|\hat{X})$.
    (b) Find a lower bound on the length $l$ of the encoded vector $Y$.
12) Determine if the following functions are convex.
    (a) $f(x) = \sqrt{x}$ over $(0, +\infty)$
    (b) $f(x) = -\log x$ over $(0, +\infty)$
13) For which range of values of $l \in \mathbb{R}$ is the function $f(x) = x^l$ convex over $(0, +\infty)$ ?
14) Show that if $X$ and $Y$ are independent then $H(X, Y) = H(X) + H(Y)$.
15) Let $X$,$Y$ be such that $H(X|Y) = 0$. Show that for each $y$ with $p(y) > 0$ there exists an $x$ such that $p(x|y) = 1$. Hint : Similar to Question 4 of Tutorial 1.
    *Remark:* If $H(X|Y) = 0$, the value of $Y$ determines the value of $X$ with probability 1.
16) Let $X$,$Y$,$Z$ be random variables taking values in $\mathbb{F}_2 = \{0,1\}$. Let $P(X = 1) = p$, $Z$ be independent of $X$ and $P(Z = 1) = \epsilon$ , and let $Y = X\ XOR\ Z$. ($a\ XOR\ b = 1$ if $a \neq b$, and $a\ XOR\ b = 0$ if $a = b$)
    (a) Find $I(X;Y)$ in terms of $\epsilon$ and $p$.
    (b) Given an $\epsilon$, for what choice of $p$ is $I(X;Y)$ maximized?
17) Suppose $X$ is a discrete random variable with alphabet $X = \{x_1, ..., x_5\}$ and probability mass function $p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.25$, $p_4 = 0.15$, $p_5 = 0.3$, where $p_i = P(X = x_i)$. Assume that you are allowed to use only $l = 2$ bits to encode/compress exactly one instance of $X$, i.e., the source encoder is a function $p : X \to \{00, 01, 10, 11\}$ that maps the information symbol $X$ into binary digits, and the source decoder is a function $\psi : \{00, 01, 10, 11\} \to X$ that estimates $X$ given the binary digits.
    (a) Consider the map $p(x_5) = 00$, $p(x_3) = 01$, $p(x_2) = 10$, $p(x_1)$=$p(x_4) = 11$. Find $P(X = x_i|p(X) = \mathbf{b})$, where $\mathbf{b} \in \{00, 01, 10, 11\}$ and $i = 1, 2, 3, 4, 5$.
    (b) Design a decoder $\psi$ for the above encoder such that
    $$\psi(\mathbf{b}) = arg\,max_{x \in X} P(X = x|p(X) = \mathbf{b}).$$
    (c) Find the probability of error $P[X \neq \psi(p(X))]$ for this encoder-decoder pair.
18) Answer briefly and directly (no need for proofs):
    (a) What are the upper and lower bounds on the probability mass function $p(x_1, ..., x_n)$ of a typical sequence $x_1, ..., x_n$?
    (b) If $n$ is sufficiently large, what are the upper and lower bounds on the size of the typical set $A_\epsilon^{(n)}$?
    (c) Is $I(X;Y)$ concave/convex with respect to $p(x)$? How about with respect to $p(y|x)$?
    (d) Write the expression for Chebyshev's inequality.
19) Prove Fano's inequality. Provide reasons for each step of the proof.
20) Show that $H(X)$ is concave in $p(x)$.
21) *Typical set for uniform distribution*
    (a) Suppose $X_1, ..., X_n$ are independent and uniformly distributed over the alphabet $X = \{0,1\}$. Find the entropy of $X_i$, and the probability mass function $p(x_1, ..., x_n)$.
    (b) Let $\epsilon > 0$. Find the exact value of the size of the typical set $A_\epsilon^{(n)}$.
22) *Stronger version of typicality.* Assume $X_1, ..., X_n$ are independent and identically distributed on a finite alphabet $X$ with the same distribution as the random variable $X$. Suppose $p(x) > 0$ for all $x \in X$.
    (a) If $1\{\cdot\}$ denotes the indicator function, find the mean of the random variable $1\{X =$

$x\}$, where $x \in X$ is a fixed constant.

(b) Consider the sequence of random variables $1\{X_1 = x\}, 1\{X_2 = x\}, ..., 1\{X_n = x\}$. Using weak law of large numbers, find the limit in probability, as $n \to \infty$, of

$$T_n(x) = \frac{1\{X_1 = x\} + ... + 1\{X_n = x\}}{n}$$

Note that $T_n(x)$ is the fraction of occurrences of $x$ is the random sequence $X_1, ..., X_n$. Also $T_n(x)$ is a random variable and is a function of the sequence $(X_1, ..., X_n)$.

(c) Let $\epsilon > 0$. Define a random variable $E_n$ as a function of $X_1, ..., X_n$ as follows. Let $E_n = 1$ if there exists at least one choice of $x \in X$ for which $|T_n(x) - p(x)| \geq \epsilon$, and $E_n = 0$ otherwise. That is, the random variable $E_n = 0$ if and only if $|T_n(x) - p(x)| < \epsilon$ for all $x \in X$ (in other words, $E_n = 0$ if and only if for every $x \in X$ , the fraction of occurrences of $x$ in $X_1, ..., X_n$ is close to $p(x)$, and $E_n = 1$ if there is even one value $x$ for which $T_n(x)$ is not close to $p(x)$).
Find the value of the limit $\lim_{n\to\infty} P(E_n = 1)$.
Hint: Express the event $E_n = 1$ as a union of $|X|$ different events $A_1, ..., A_{|X|}$. Then use the union bound on probability: $P(A_1 \cup ... \cup A_{|X|}) \leq P(A_1) + ... + P(A_{|X|})$. Use weak law of large numbers to upper bound $\lim_{n\to\infty} P(A_i)$, and use this to upper bound $\lim_{n\to\infty} P(E_n = 1)$.

23) Consider $n$ pairs of random variables $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$, where each pair $(X_i, Y_i)$ is distributed as $p(x, y)$ and the $n$ pairs are independent of each other. Let

$$Z_n = \frac{1}{n} \log \frac{p(X_1, Y_1, X_2, Y_2, ..., X_n, Y_n)}{p(X_1, X_2, ..., X_n)}.$$

Find the limit (in probability) of the sequence of random variables $Z_1, Z_2, ..., Z_n$...

24) Let $X$ be a random variable distributed on a finite alphabet $X$ with probability mass function $p(x)$. A random variable $Y = f(X)$, with a finite alphabet $Y$, is obtained by applying a function $f : X \to Y$ on $X$. Finally, the function $g : Y \to X$ produces an estimate $\hat{X} = g(Y)$ of the random variable $X$.

$$X \longrightarrow \boxed{f} \xrightarrow{Y} \boxed{g} \longrightarrow \hat{X}$$

(a) If $P(\hat{X} \neq X) = \epsilon$, derive the following lower bound on $\log |Y|$ in terms of $\epsilon$, $H(X)$ and $|X|$ : $\log |Y| \geq H(X) - 1 - \epsilon \log |X|$.

(b) Suppose $X = (Z_1, ..., Z_n)$, where each $Z_i$ is independent and identically distributed over the alphabet $Z$ as $p(z)$, and say $Y = \{0, 1\}^m$ is the set of all binary vectors of length $m$. Also assume that $m = nR$, where $R$ is a constant. Show that if $R < H(Z)$, as $n \to \infty$, $\epsilon$ does not converge to 0 irrespective of the functions $f$ and $g$. (For simplicity assume that $m = nR$ is an integer for every $n$) *Remark: Here, $f$ is the source encoder, $g$ is the decoder, and $R$ is the rate of the code. Note that this a fixed length codebook, i.e., all codewords have the same length $m$. This is different from variable length prefix-free source codes discussed in class. Unlike the problem scenario discusses in class, here we are allowed to make decoding errors. The probability of these errors $\epsilon$ must be small for a good system. The above problem shows that for $\epsilon$ to be small, we necessarily require $R \geq H(X)$, i.e., the codelength $m$ should scale as $H(X) \times n$ or faster with blocklength $n$.*

25) Let $X_1, ..., X_n$ be independent and identically distributed binary random variables with $P(X_i = 1) = \frac{1}{4}$ and $P(X_i = 0) = \frac{3}{4}$.

(a) Find the entropy $H(X)$ of the random variable with $P(X = 0) = \frac{3}{4}$ and $P(X = 1) = \frac{1}{4}$.

(b) Let $t$ be the number of ones in the binary sequence $(x_1, ..., x_n)$. Express $p(x_1, ..., x_n)$, i.e.,$P((X_1, ..., X_n) = (x_1, ..., x_n))$ in terms of $t$ and $n$.

(c) For $\epsilon > 0$, consider the typical set $A_\epsilon^{(n)}$. Use the upper and lower bounds on $p(x_1, ..., x_n)$ for a typical sequence $x_1, ..., x_n$ to show the following: a binary vector $(x_1, ..., x_n)$ belongs to the typical set $A_\epsilon^{(n)}$ if and only if the number of ones $(1's)$ in $\overline{(x_1, ..., x_n)}$ lies between

$$\frac{n}{4} - \frac{n\epsilon}{\log 3} \ and \ \frac{n}{4} + \frac{n\epsilon}{\log 3}.$$

(d) Using the above results, prove the following: for any integer $n$ that is multiple of $4$, the binomial coefficient $\binom{n}{\frac{n}{4}}$ satisfies

$$\frac{1}{n} \log \left[ \binom{n}{\frac{n}{4}} \right] \leq H(X) + \epsilon.$$

*Hint:* Use the upper bound on the size of the typical set.

*Remark: We can conclude that* $\frac{1}{n} \log \left[ \binom{n}{\frac{n}{4}} \right] \leq H(X)$*, since the above inequality holds for any* $\epsilon > 0$*.*

## REFERENCES


[1] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1999.

[2] P. V. Kumar. [Online]. Available: http://ece.iisc.ernet.in/~vijay/it.html)