MTHE 474 Notes

Timothy Liu

Fall 2022

Contents

1 Chapter 2

1.1 Information Measures for Discrete Systems

1.1.1 Definitions

- **Definition 2.2:** Entropy of discrete random variable X with pmf $P_X(*)$ is defined as

$$H(x) := -\sum_{x \in X} P_X(x) * \log_2 P_X(x)$$

- **Definition 2.2:** Entropy of discrete random variable X with pmf $P_X(*)$ is defined as

$$H(x) := -\sum_{x \in X} P_X(x) * \log_2 P_X(x)$$

- **Definition 2.8 (Joint entropy):**

$$H(X,Y) := -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x,y) * \log_2 P_(X,Y)(x,y)$$

- **Definition 2.9 (Conditional entropy):**

$$H(Y|X) := \sum_{x \in \mathcal{X}} P_X(x)(-\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) * \log_2 P_{Y|X}(y|x))$$

-

1.1.2 Lemmas/Theorems

- **Lemma 2.4 (Fundamental Inequality):** $\forall\ x > 0$ and $D > 1$ we have

$$\log_D(X) \leq \log_D e * (x - 1)$$

- **Lemma 2.5 (Non-negativity):** $H(X) \geq 0$
- **Lemma 2.6 (Entropy Upper-Bound):** $H(X) \leq \log_2 |\mathcal{X}|$ where random variable X takes values from finite set $\mathcal{X}$
- **Lemma 2.7 (Log-Sum inequality)** Write this one out later
- **Theorem 2.10 (Chain rule for entropy):** $H(X,Y) = H(X) + H(Y|X)$

1.2 Mutual Information

1.2.1 Definitions

- **Definition 2.2.1 (Mutual Information):**

$$I(X;Y) := H(X) - H(X|Y)$$

- **Definition 2.2.2 (Conditional Mutual Information):**

$$I(X;Y|Z) := H(X|Z) - H(X|Y,Z)$$

1.2.2 Lemmas

- **Lemma 2.15 (Properties of Mutual Information):**

$$1.\ I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log_2 \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \tag{1}$$

$$2.\ I(X;Y) = I(Y;X) = H(Y) - H(Y|X) \tag{2}$$

$$3.\ I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{3}$$

$$4.\ I(X;Y) \leq H(X) \quad \text{equality iff X is a function of Y} \tag{4}$$

$$5.\ I(X;Y) \leq 0 \text{ with equality iff X and Y are independent} \tag{5}$$

$$6.\ I(X;Y) \leq \min\{\log_2 |\mathcal{X}|, \log_2 |\mathcal{Y}|\} \tag{6}$$

- **Lemma 2.16 (Chain Rule for Mutual Information):**

$$I(X;Y,Z) = I(X;Y) + I(X;Z|Y) = I(X;Z) + I(X;Y|Z)$$

- **Theorem 2.17 (Chain Rule for entropy):** $X^n := (X_1, \ldots, X_n)$ and $x^n := (x_1, \ldots, x_n)$

$$H(X^n) = \sum_{i=1}^{n} H(X_i|X^{i-1})$$

- **Theorem 2.18 (Chain Rule for conditional entropy):**

$$H(X_1, X_2, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1, Y)$$

- **Theorem 2.19 (Chain Rule for Mutual information):**

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \ldots, X_1)$$

Where $I(X_i; Y|X_{i-1}, \ldots, X_1) := I(X_1, Y)$ for $i = 1$

1.3 Conditional Divergence

1.3.1 Definitions

1.3.2 Theorems

1.4 Data Processing Inequality

1.4.1 Definitions

- **Lecture 7 Definition (Markov Chain):** Three jointly distributed random variables X, Y, Z are said to form a Markov Chain (in that order), denoted by $X \to Y \to Z$
if:

$$P_{XZ|Y}(x, y|z) = P_{X|Y}(x|y)P_{Z|Y}(z, y) \Longleftrightarrow P_{Z|XY}(z|x, y) = P_{Z|Y}(z|y)$$

$\forall x \in X, y \in Y, z \in Z$

1.4.2 Theorems

- **Lecture 7 Theorem (Data Processing Inequality):** If $X \to Y \to Z$, then

$$I(X;Y) \leq I(X;Z)$$

1.5 Convex/Concavity of Information Measures

1.5.1 Definitions

- **Lecture 6 Definition (Convex Set):**

A subset K of $\mathbb{R}$ is called convex if the line segment joining any two points in K also lies in K

- **Lecture 6 Definition (Convex Function):** The function $f : k \to \mathbb{R}$ where k is a convex subset of $\mathbb{R}^n$, is called convex on k if $\forall x_1, x_2 \in k$ and $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

Strict equality holds whenever $x_1 \neq x_2$ and $0 < \lambda < 1$ then f is called strictly convex

- **Lecture 6 Definition (Concave Function):** $f : k \to \mathbb{R}$ is concave on k (where $k \subseteq \mathbb{R}^n$ is a concave subset) if $-f$ is convex. In other words: if $\forall x_1, x_2 \in k$ and $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

1.5.2 Theorems

- **Lecture 6 Theorem (Jensen's Inequality):** Let $K \subseteq \mathbb{R}$ and let $f : k \to \mathbb{R}$ be a convex function. Also let x be a RV with alphabet $\mathcal{X} \subseteq k$ and finite mean, then

$$E[f(x)] \leq f(E[x])$$

  Also if f is strictly convex, then the inequality is strict unless x is deterministic

- **Lecture 7 Theorem (Convexity/Concavity of Information Measures):**
  i. $D(p||q)$ is convex in the pair (p, q) (ie: if p_1, q_1 and p_2, q_2 are two pairs of PMFs defined on $\mathcal{X}$) then:

$$D(\Lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2) \leq \lambda d(p_1||q_1) + (1-\lambda)D(p_2||q_2)$$

  $\forall \lambda \in [0, 1]$

  ii. if x P_x, then

$$H(x) = H(p_x) \text{ is concave in } P_x$$

  iii. If (x, y) $P_X P_{Y|X}$, then $I(X; Y) = I(P_X, P_{Y|X})$ is concave in P_X for fixed $P_{Y|X}$ and convex in $P_{Y|X}$ for fixed P_X