信息论复习问题集（含答案）

基本定义复述

1. 请严格复述**熵 (Entropy)** 的定义（包含数学表达式和定义域说明）。

> **答案**：离散随机变量 X 取值于有限集 $\mathcal{X}$，其熵定义为:
>
> $$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$$
>
> 其中约定 $0\log 0 = 0$，对数底数通常为 2（单位为比特）或 e（单位为奈特）。

2. 请严格复述**联合熵 (Joint Entropy)** 的定义（包含数学表达式）。

> **答案**：随机变量 (X,Y) 服从联合分布 $p(x,y)$，其联合熵定义为:
>
> $$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y)$$

3. 请严格复述**条件熵 (Conditional Entropy)** 的定义（包含两种等价表达式）。

答案：给定随机变量 Y 时 X 的条件熵定义为：

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x|y)$$

4. 请严格复述**相对熵 (Kullback-Leibler Divergence)** 的定义（包含数学表达式和非负性说明）。

答案：对同支撑集的两个概率分布 $p(x)$ 和 $q(x)$，相对熵定义为：

$$D(p \parallel q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$$

满足 $D(p \parallel q) \geq 0$，等号成立当且仅当 $p = q$（由 Jensen 不等式证明）。

5. 请严格复述**互信息 (Mutual Information)** 的定义（包含三种等价表达式）。

答案：随机变量 X 和 Y 的互信息定义为：

$$I(X;Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = D(p(x,y) \parallel p(x)p(y))$$

等价形式：

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

定理证明

6. **链式法则 (Chain Rules)**

(a) 请证明**熵的链式法则**：

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

证明： 由条件熵定义展开：

$$\begin{aligned}
H(X_1, \ldots, X_n) &= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n) \\
&= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log \prod_{i=1}^{n} p(x_i|x_1, \ldots, x_{i-1}) \\
&= -\sum_{i=1}^{n} \sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_i|x_1, \ldots, x_{i-1}) \\
&= \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}) \quad \square
\end{aligned}$$

(b) 请证明**互信息的链式法则**：

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

证明：由互信息和条件互信息定义：

$$\begin{aligned} I(X_1,\ldots,X_n;Y) &= H(X_1,\ldots,X_n) - H(X_1,\ldots,X_n|Y) \\ &= \sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1}) - \sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1},Y) \\ &= \sum_{i=1}^{n} [H(X_i|X_1,\ldots,X_{i-1}) - H(X_i|X_1,\ldots,X_{i-1},Y)] \\ &= \sum_{i=1}^{n} I(X_i;Y|X_1,\ldots,X_{i-1}) \quad \square \end{aligned}$$

7. **琴生不等式 (Jensen's Inequality)**：请证明对于凸函数 f 和随机变量 X：

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

证明：设 f 为凸函数，则对任意 $\lambda \in [0,1]$ 和 x_1, x_2 满足：

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

由数学归纳法可推广到有限点：对概率分布 p_i 和点 x_i，

$$f\left(\sum p_i x_i\right) \leq \sum p_i f(x_i)$$

即 $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$。当 f 严格凸时，等号成立当且仅当 X 为常数。
$\square$

8. **马尔科夫链 (Markov Chain)**：

   (a) 请给出 $X \to Y \to Z$ 构成马尔科夫链的**严格概率定义**

答案： $X \to Y \to Z$ 构成马尔科夫链当且仅当满足：

$$p(x,y,z) = p(x) \cdot p(y|x) \cdot p(z|y)$$

等价于条件独立：$p(z|x,y) = p(z|y)$

(b) 请证明马尔科夫链的等价条件：$I(X;Z|Y) = 0$

证明：

$$\begin{aligned} I(X;Z|Y) &= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,z|y)}{p(x|y)p(z|y)} \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(z|x,y)p(x|y)}{p(x|y)p(z|y)} \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(z|y)}{p(z|y)} \quad (\text{由马尔科夫性} p(z|x,y) = p(z|y)) \\ &= 0 \quad \square \end{aligned}$$

9. **数据处理不等式 (Data Processing Inequality)**：若 $X \to Y \to Z$ 形成马尔科夫链，请证明：

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

证明： 由互信息链式法则和马尔科夫性：

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

由马尔科夫性 $I(X;Z|Y) = 0$，且 $I(X;Y|Z) \geq 0$，故：

$$I(X;Y) = I(X;Z) + I(X;Y|Z) \geq I(X;Z) \quad \square$$

10. **法诺不等式 (Fano's Inequality)**：

(a) 请写出完整的法诺不等式表达式

答案：对满足 $X \to Y \to \hat{X}$ 的估计量 $\hat{X}$，错误概率 $P_e = \Pr(X \neq \hat{X})$：

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|\hat{X}) \geq H(X|Y)$$

其中 $H(P_e) = -P_e \log P_e - (1 - P_e) \log(1 - P_e)$ 为二元熵。

(b) 请解释 P_e、$H(P_e)$、$H(X|\hat{X})$ 和 $H(X|Y)$ 的含义

答案：

- P_e：估计错误概率 $\Pr(X \neq \hat{X})$
- $H(P_e)$：二元熵，表示错误事件的不确定性
- $H(X|\hat{X})$：已知估计值 $\hat{X}$ 后 X 的条件熵
- $H(X|Y)$：已知观测 Y 后 X 的最小可能条件熵

(c) 请说明马尔科夫链条件 $X \to Y \to \hat{X}$ 的重要性

答案：该条件确保 $\hat{X}$ 仅通过 Y 依赖于 X（即无额外信息），使得 $H(X|Y)$ 是理论最小不确定性，且不等式 $H(X|\hat{X}) \geq H(X|Y)$ 成立。

11. **条件作用使熵减少**：请证明对于任意随机变量 X 和 Y：

$$H(X|Y) \leq H(X)$$

证明：由互信息非负性：

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \geq 0 \\ \Rightarrow H(X|Y) &\leq H(X) \end{aligned}$$

等号成立当且仅当 X 与 Y 独立。□

12. **熵的独立界放缩**：请证明若随机变量 $X_1, X_2, \ldots, X_n$ 相互独立，则：

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

证明：由独立性和联合熵定义：

$$\begin{aligned} H(X_1, \ldots, X_n) &= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log \prod_{i=1}^{n} p(x_i) \\ &= -\sum_{i=1}^{n} \sum_{x_i} p(x_i) \log p(x_i) \quad \text{(由独立性)} \\ &= \sum_{i=1}^{n} H(X_i) \quad \square \end{aligned}$$