1.(每题 4 分，共 8 分) 选择题 (多选题)

(a) 设 X,Y,Z 为离散随机变量，U,V 为连续随机变量，则以下等式或不等式正确的是 (AB)

$$(A)H(g(X)) \leqslant H(X)$$

$$(B)若X,Y独立，则有H(XY)=H(X)+H(Y)$$

$$(C)h(U) \geqslant I(U;V)$$

$$(D)I(X;Y|Z) \leqslant I(X;Y)$$

解析：

(A) 正确. 为了比较二者的关系，我们首先要在二者之间建立联系，即考察二者的联合熵 $H(X,g(X))$. 我们对联合熵 $H(X,g(X))$ 尝试**两种不同方式的分解**，进而可以建立 $H(X)$ 和 $H(g(X))$ 的联系.

第一种分解

$$\begin{aligned} H(X,g(X)) &\stackrel{(a)}{=} H(X)+H(g(X)\mid X) \\ &\stackrel{(b)}{=} H(X) \end{aligned}$$

第二种分解

$$\begin{aligned} H(X,g(X)) &\stackrel{(c)}{=} H(g(X))+H(X\mid g(X)) \\ &\geqslant H(g(X)) \end{aligned}$$

其中

(a) 由熵的链式法则可知.

(b) 由于给定了 $X=x$ 后，$g(X)=g(x)$ 为定值，所以条件熵为 0.

(c) 由熵的链式法则可知.

(d) 由于条件熵非负. 等号成立条件为：$g(x)$ 为可逆映射.

由上述讨论可知

$$H(g(X)) \leqslant H(X)$$

可得结论正确.

(B) 正确. 设两个统计独立的信源 X 和 Y 的信源空间分别为

$$[X \cdot P]: \begin{cases} X: & a_1 \quad a_2 \quad \cdots \quad a_r \\ P(X): & p_1 \quad p_2 \quad \cdots \quad p_r \end{cases}$$

其中 $0 \leqslant p_i \leqslant 1(i=1,2,\ldots,r)$，$\sum\limits_{i=1}^{r} p_i = 1$.

$$[Y \cdot P]:\left\{\begin{array}{ccccc} Y: & b_1 & b_2 & \cdots & b_s \\ P(Y): & q_1 & q_2 & \cdots & q_s \end{array}\right.$$

其中 $0 \leqslant q_j \leqslant 1(j=1,2,\ldots,s)$，$\sum\limits_{i=1}^{s} q_j = 1$.
若信源 X 和 Y 同时发出各自的符号，则构成联合信源 (XY)，其符号表为

$$(XY):\{(a_i b_j)\,(i=1,2,\ldots,r; j=1,2,\ldots,s)\}$$

概率分布为

$$p\left(a_i b_j\right)=p\left(a_i\right) p\left(b_j\right)=p_i q_j \quad (i=1,2,\ldots,r; j=1,2,\ldots,s)$$

满足

$$\sum_{i=1}^{r} \sum_{j=1}^{s} p\left(a_i b_j\right)=\sum_{i=1}^{r} \sum_{j=1}^{s} p\left(a_i\right) p\left(b_j\right)=\sum_{i=1}^{r} \sum_{j=1}^{s} p_i q_j=1$$

所以有

$$\begin{aligned}
H(XY) &= -\sum_{i=1}^{r} \sum_{j=1}^{s} p\left(a_i b_j\right) \log p\left(a_i b_j\right) \\
&= -\sum_{i=1}^{r} \sum_{j=1}^{s} p_i q_j \log p_i q_j \\
&= -\sum_{i=1}^{r} \sum_{j=1}^{s} p_i q_j \log p_i - \sum_{i=1}^{r} \sum_{j=1}^{s} p_i q_j \log q_j \\
&= \sum_{j=1}^{s} q_j \left\{-\sum_{i=1}^{r} p_i \log p_i\right\} + \sum_{i=1}^{r} p_i \left\{-\sum_{j=1}^{s} q_j \log q_j\right\} \\
&= \sum_{j=1}^{s} q_j H\left(p_1, p_2, \ldots, p_r\right) + \sum_{i=1}^{r} p_i H\left(q_1, q_2, \ldots, q_s\right) \\
&= \sum_{j=1}^{s} q_j H(X) + \sum_{i=1}^{r} p_i H(Y) \\
&= H(X) + H(Y)
\end{aligned}$$

所以命题成立. 实际上，这个命题叫做熵函数可加性定理. 熵函数的可加性表明, 由两个相互统计独立的离散信源 X 和 Y 构成的联合信源 (XY) 的熵，等于信源 X 和 Y 各自熵之和. 联合信源 (XY) 每发一个消息所能提供的平均信息量, 等于信源 X 和 Y

各自发一个符号提供的平均信息量之和 (当 X 和 Y 不统计独立, 相互间存在统计依赖关系时, 熵函数的可加性同样成立). 熵函数的可加性的数学意义是：**两个熵函数之和仍然是一个熵函数**.

(C) 错误. 反例：U 为 $[0,\frac{1}{2}]$ 上的均匀分布，$h(U)=\log\frac{1}{2}=-1$ 比特，对于 V 服从任意分布，都有 $I(U;V)\geqslant 0$. 所以命题错误.

(D) 错误." 条件使得互信息减少" 这一结论是有成立条件的，要求 $X\to Y\to Z$ 构成马尔可夫链. 若对于一般的情形，存在 X,Y,Z 使得 $I(X;Y|Z)\geqslant I(X;Y)$.

考虑 X,Y 独立情况下，$I(X;Y)=0$. 记 $Z=X+Y$，有

$$I(X;Y\mid Z)=H(X\mid Z)-H(X\mid Y,Z)=H(X\mid Z)\geqslant 0$$

若 Y 不是常量，则可以进一步得 $I(X;Y\mid Z)>I(X;Y)=0$.

(b) 以下 D 元字母表上的码字长度符合即时码要求的是 (ACD)

$$
\begin{aligned}
&(A)D=2,l_i=2,3,3,4,4,4,\cdots,\overbrace{i,i\cdots,i}^{i-1},\cdots\\
&(B)D=2,l_i=1,3,3,3,4,5,6,6,6\\
&(C)D=3,l_i=1,2,2,3,3,3,4,4,5,5,5,6,7,8,8,8,9,9,9,9,9\\
&(D)D=4,l_i=1,1,1,2,2,3,3,3,4
\end{aligned}
$$

解析：

(A) 记 $S=\sum_{k=2}^{n}2^{-k}(k-1)$，乘以 1/2 后错位相减可得

$$S-\frac{1}{2}S=\sum_{k=2}^{n}2^{-k}-\frac{n-1}{2^{n+1}}$$

即

$$S=1-\frac{n+1}{2^n}\leqslant 1$$

所以符合即时码要求. (B)

$$2^{-1}+2^{-3}\times 3+2^{-4}+2^{-5}+2^{-6}\times 3>2^{-1}+2^{-3}\times 3+2^{-4}+2^{-4}=2^{-1}+2^{-3}\times 4=1$$

(C)

$$3^{-1}+3^{-2}\times 2+3^{-3}\times 3+...<3^{-1}+3^{-2}\times 2+3^{-3}\times 3+<3^{-4}\times 27=1$$

(D)

$$4^{-1}\times 3+4^{-2}\times 2+4^{-3}\times 3+4^{-4}<4^{-1}\times 3+4^{-2}\times 3+4^{-3}\times 4=1$$

2.(每题 4 分，共 8 分) 判断题 (若判断为对，简要说明或证明；若判断为错，简要说明或举出反例)

(a) 假设随机变量 X，Y 和 Z,W 构成如下的马尔科夫链

$$X \rightarrow Y \rightarrow (Z,W),\ 即 p(x,y,z,w)=p(x)p(y\mid x)p(z,w\mid y)$$

则有

$$I(X;Z)+I(X;W)\leqslant I(X;Y)+I(Z;W)$$

(b) 给定一个概率分布 $(p_1,p_2,\ldots,p_n)$ 和一个整数 $m(0\leqslant m\leqslant n)$. 定义 $q_m=1-\sum_{i=1}^{m}p_i$，则有

$$H\left(p_1,p_2,\ldots,p_n\right)\leqslant H\left(p_1,p_2,\ldots,p_m,q_m\right)+q_m\log(n-m)$$

(a) 对.

证明：

由数据处理不等式知

$$I(X;Y)\geqslant I(X;(Z,W))$$

于是有

$$\begin{aligned}
&I(X:Y)+I(Z;W)-I(X;Z)-I(X;W)\\
&\quad\geqslant I(X:Z,W)+I(Z;W)-I(X;Z)-I(X;W)\\
&\quad=H(Z,W)+H(X)-H(X,W,Z)+H(W)+H(Z)-H(W,Z)\\
&\quad=-H(X,W,Z)+H(X,Z)+H(X,W)-H(X)\\
&\quad=H(W\mid X)-H(W\mid X,Z)\\
&\quad=I(W;Z\mid X)\\
&\quad\geqslant 0
\end{aligned}$$

(b) 对.

证明：

要证明

$$H\left(p_1,p_2,\ldots,p_n\right)\leqslant H\left(p_1,p_2,\ldots,p_m,q_m\right)+q_m\log(n-m)$$

即证明

$$-\sum_{i=1}^{n}p_i\log p_i\leqslant-\sum_{i=1}^{m}p_i\log p_i+q_m\log q_m+q_m\log(n-m)$$

即只需证明

$$-\sum_{i=m+1}^{n} p_i \log p_i \leqslant q_m \log q_m + q_m \log(n-m)$$

构造概率分布

$$(r_1, r_2, \ldots, r_{n-m}) = \left(\frac{p_{m+1}}{q_m}, \ldots, \frac{p_n}{q_m}\right)$$

由熵的极值性可知，当 $r_1 = r_2 = \cdots = r_{n-m}$ 时，有

$$H(r_1, r_2, \ldots, r_{n-m}) \leqslant \log|n-m|$$

即

$$-\sum_{i=m+1}^{n} \frac{p_i}{q_m} \log \frac{p_i}{q_m} \leqslant \log|n-m|$$

即

$$-\sum_{i=m+1}^{n} p_i \log p_i + \sum_{i=m+1}^{n} \log q_m \leqslant q_m \log|n-m|$$

即

$$-\sum_{i=m+1}^{n} p_i \log p_i \leqslant q_m \log q_m + q_m \log(n-m)$$

所以命题得证，即

$$H(p_1, p_2, \ldots, p_n) \leqslant H(p_1, p_2, \ldots, p_m, q_m) + q_m \log(n-m)$$

其中等号成立条件为

$$p_{m+1} = p_{m+2} = \ldots = p_n$$

3.(每题 4 分，共 16 分) 填空题

(a) 有三个二元离散随机变量 X, Y, Z，若要使得 $I(X;Y) = 1$ bit ，$I(X;Y \mid Z) = 0$ bit，则 X, Y, Z 的联合概率分布为_________ .

(b) 给定连续随机变量 X、Y，服从分布 $\mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, 0)$. 设 $U = X + Y$，$V = X - Y$，则互信息 $I(U;V) =$_________ .

(c) 考虑随机变量 X，取 6 个值 $\{A, B, C, D, E, F\}$ ，其概率依次为 0.5，0.25 ，0.1，0.05，0.05 和 0.05. 构造该随机变量的四元哈夫曼码 (字母表由 4 个元素构成，例如 a, b, c, d). 其期望长度为_________ .

(d) 设一有线电话信道，带宽限制在 $200 \sim 3600Hz$，信噪比为 $\dfrac{P}{N_0} = 1000$，则该信道的信道容量为_________

答案：(a)$p(X = i, Y = j, Z = k) = p(X = 1 - i, Y = 1 - j, Z = 1 - k) = \dfrac{1}{2}$.

(b)$\log\dfrac{\sigma_x^2+\sigma_y^2}{\sigma_x\sigma_y}-1$.(c)1.15 四进制信息单位.(d)381 比特/秒.

解析：

(a) 由于

$$I(X;Y)=H(X)-H(X\mid Y)\le H(X)\le\log|\mathcal{X}|=1\text{比特}$$

所以可知，等号取到，则有

$$H(X)=1\text{比特}\quad H(X\mid Y)=0$$

即 X 为均匀分布 ($p(X=0)=p(X=1)=\frac{1}{2}$)，且给定 Y 时 X 确定，对于二元分布即 $Y=X$ 或者 $Y=-X$. 第二个互信息等式，有互信息为 0 的意义可知，给定 Z 的时候 X,Y 独立. 但是我们由第一个等式分析得到，二者给定一个另一个就唯一确定，那么独立只可能是一种情形，即给定 Z 的时候，X,Y 都为确定量 $(p(x,y\mid z)=p(x\mid z)p(y\mid z)=1)$.

综上所述，可得满足题意的概率分布为

$$p(X=i,Y=j,Z=k)=p(X=1-i,Y=1-j,Z=1-k)=\frac{1}{2}$$

(b) 首先根据联合概率密度函数可以分析得到 $X\sim\mathcal{N}(m_x,\sigma_x^2)$，$Y\sim\mathcal{N}(m_y,\sigma_y^2)$，$X,Y$ 独立. 则可知 $U=X+Y\sim\mathcal{N}(m_x+m_y,\sigma_x^2+\sigma_y^2)$，$V=X-Y\sim\mathcal{N}(m_x-m_y,\sigma_x^2+\sigma_y^2)$. 特别要注意的是，正态分布的差的分布中，均值为二者均值之差，方差为二者之和.

于是可得

$$h(U)=h(V)=\frac{1}{2}\left[\log 2\pi e(\sigma_x^2+\sigma_y^2)\right]\text{ 比特}$$

利用互信息的分解式

$$I(U;V)=h(U)+h(V)-h(U,V)=\left[\log 2\pi e(\sigma_x^2+\sigma_y^2)\right]-h(U,V)$$

这时我们需要用到线性变换对于微分熵的影响这一性质. 我们在前面有介绍过

$$h(A\boldsymbol{X})=h(\boldsymbol{X})+\log|A|$$

对于本题，有

$$\begin{pmatrix}U\\V\end{pmatrix}=\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}X\\Y\end{pmatrix}$$

其中

$$A=\begin{pmatrix}1&1\\1&-1\end{pmatrix}$$

其行列式为 $|A| = 2$. 进而可得

$$h(U,V) = h(X,Y) + \log 2 = h(X) + h(Y) + 1 = \frac{1}{2}\left[\log 2\pi e(\sigma_x^2)\right] + \frac{1}{2}\left[\log 2\pi e(\sigma_y^2)\right] + 1\text{比特}$$

所以可得，互信息为

$$\begin{aligned} I(U;V) &= h(U) + h(V) - h(U,V) \\ &= \left[\log 2\pi e(\sigma_x^2 + \sigma_y^2)\right] - \left[\frac{1}{2}\left[\log 2\pi e(\sigma_x^2)\right] + \frac{1}{2}\left[\log 2\pi e(\sigma_y^2)\right] + 1\right] \\ &= \log \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x \sigma_y} - 1 \quad \text{比特} \end{aligned}$$

(c) 四元哈夫曼编码，首先检查是否需要添加零概率项.

$$|\mathcal{X}| = 6 = 1 + k(4-1)$$

解得 $1 < k = 5/3 < 2$. 即无整数解 k. 所以需要添加零概率项. 零概率项个数为

$$2 \times (4-1) + 1 - 6 = 1$$

加入 1 个零概率项 G 后构造四元哈夫曼码树进行编码.

码字		概率		
a	A	0.5	0.5	1.0
b	B	0.25	0.25	
d	C	0.1	0.15	
ca	D	0.05	0.1	
cb	E	0.05		
cc	F	0.05		
cd	G	0.0		

期望码长为

$$1 \times (0.5 + 0.25 + 0.1) + 2 \times (0.05 + 0.05 + 0.05) = 1.15 \quad \text{四进制信息单位}$$

(d) 有限带宽信道容量计算公式

$$C = W \log\left(1 + \frac{P}{N_0 W}\right) \text{ 比特/ 秒}$$

由题意知，$W=3600-200=3400Hz$，$P/N_0=1000$. 代入可得

$$C=3400\log\left(1+\frac{1000}{3400}\right)=381 \text{ 比特/ 秒}$$

4.(10 分) 随机变量 X,Y 的取值分别为 $x_1,x_2,\ldots,x_m$ 和 $y_1,y_2,\ldots,y_n$. 设 $Z=X+Y$.

(1) 如果 X、Y 独立，证明 $H(Y)\leqslant H(Z)$ 和 $H(X)\leqslant H(Z)$.

(2) 请给出一个例子，使得 $H(X)>H(Z)$ 且 $H(Y)>H(Z)$.

(3) 说明在什么条件下，熵的和等于和的熵，即 $H(Z)=H(X)+H(Y)$.

解析：

(1) 这两条结论是对称的，只需要讨论其中一个. 我们不妨讨论第一条结论. 为了证明这条结论，我们要考虑如何将 $H(Z)=H(X+Y)$ 和 $H(Y)$ 建立联系. 根据题目中给的条件，由 X,Y 独立，有 $H(Y)=H(Y\mid X)$. 这样，利用这一条件，待证明的结论可以等价写作 $H(Y\mid X)\leqslant H(Z)$. 由随机变量熵的性质，我们考虑使用“条件使得熵减少”来完成不等式的放缩. 由此出发，我们考虑证明 $H(Y\mid X)=H(Z\mid X)$.

由 $Z=X+Y$ 知，$p(Z=z\mid X=x)=p(Y=z-x\mid X=x)$. 进而，

$$\begin{aligned}
H(Z\mid X)&=\sum p(x)H(Z\mid X=x)\\
&=-\sum_x p(x)\sum_z p(Z=z\mid X=x)\log p(Z=z\mid X=x)\\
&=\sum_x p(x)\sum_y p(Y=z-x\mid X=x)\log p(Y=z-x\mid X=x)\\
&=\sum p(x)H(Y\mid X=x)\\
&=H(Y\mid X).
\end{aligned}$$

这样，我们可得

$$H(Y)=H(Y\mid X)=H(Z\mid X)\leqslant H(Z)$$

另一条对称的结论同理可得.

(2) 这种举例子的问题，一种常用的策略是让 $Z=X+Y$ 为常量，这样由熵的非负性即可得到要满足的条件. 具体地，例如

$$X=-Y=\begin{cases}1, & p=1/2\\ 0, & p=1/2\end{cases}$$

$H(X)=H(Y)=1$ 比特，而 $H(Z)=H(X+Y)=0$ 比特.

(3) 由熵的性质可知

$$H(X,Y)=H(X,X+Y)=H(X+Y)+H(X\mid X+Y)\geqslant H(X+Y)$$

和

$$H(X,Y)=H(X)+H(Y\mid X)\leqslant H(X)+H(Y)$$

这样，我们通过 $H(X,Y)$ 建立了 $H(X+Y)$ 和 $H(X)+H(Y)$ 之间的联系

$$H(Z)\leqslant H(X,Y)\leqslant H(X)+H(Y)$$

我们考察两个不等式的取等条件，即可得到本题的答案. 对于第一个不等式，等号成立条件是 $H(X\mid X+Y)=0$，即给定了 $X+Y$ 的取值后 X 为定值，也就是 Y 为常量. 对于第二个不等式，等号成立条件是 $H(Y\mid X)=H(Y)$，即 X、Y 独立. 结合两个条件可得，$H(Z)=H(X)+H(Y)$ 的成立条件为：X、Y 中至少一个为常量.

5.(8 分) 一只小鸟在 $3\times3\times3$ 的立方体迷宫中迷失了. 这只鸟在相互邻接的房间之间，从这个房间穿过墙飞到那个房间的概率是相同的. 例如，角落边的房间有 3 个出口. 求此随机过程的熵率.

解析：

加权图上的随机移动问题. 首先建立模型，并计算 W_i 和 W.

中心：$W_1=6$.

面中心：$W_{2_j}=5(j=1,2,\ldots,6)$.

棱的中点：$W_{3_j}=4(j=1,2,\ldots,12)$.

顶点：$W_{4_j}=3(j=1,2,\ldots,8)$.

所以 $2W=\sum W_i=108$. 则可得平稳分布

$$\mu_1=\frac{6}{108}=\frac{1}{18}$$

$$\mu_{2_j}=\frac{5}{18}(j=1,2,\ldots,6)$$

$$\mu_{3_j}=\frac{1}{27}(j=1,2,\ldots,12)$$

$$\mu_{4_j}=\frac{1}{36}(j=1,2,\ldots,8)$$

即平稳分布为

$$\boldsymbol{\mu}=\left(\frac{1}{18},\frac{5}{18},\frac{5}{18},\frac{5}{18},\frac{5}{18},\frac{5}{18},\frac{5}{18},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27},\frac{1}{27}\right.$$
$$\left.,\frac{1}{36},\frac{1}{36},\frac{1}{36},\frac{1}{36},\frac{1}{36},\frac{1}{36},\frac{1}{36},\frac{1}{36}\right)$$

进而，利用结论

$$H\left(\cdots,\frac{W_{ij}}{2W},\cdots\right)-H\left(\cdots,\frac{W_i}{2W},\cdots\right)$$

结合前述分析可得

$$
\begin{aligned}
H(\mathscr{X}) &= \log(108) + 8\left(\frac{3}{108}\log\frac{3}{108}\right) + 12\left(\frac{4}{108}\log\frac{4}{108}\right) + 6\left(\frac{5}{108}\log\frac{5}{108}\right) + 1\left(\frac{6}{108}\log\frac{6}{108}\right) \\
&= 2.03\text{比特}
\end{aligned}
$$

6.(10 分) 考虑哈夫曼编码问题. 假定消息的概率分布以递减的顺序给出 $p_1 > p_2 > \cdots > p_m$.

(a) 证明：对任意的二元哈夫曼码，如果最可能出现的消息字符的概率 $p_1 > 2/5$，则该字符分配的码字长度必为 1.

(b) 对任意的二元哈夫曼码，如果最可能出现的消息字符的概率 $p_1 < 1/3$，则该字符分配的码字长度必超过 1.

证明：

(a) 反证法. 假设分配给 p_1 的码字长度 $l_1 \geqslant 2$. 设 p_1 对应的码字以 00 开头，并以 C_x 表示以 x 开头的码字，由题意知 C_{00} 出现的概率大于 2/5，则 C_{01}, C_{10}, C_{11}(按出现概率由高到低排列) 出现的概率和小于 3/5. 其中 C_{10}, C_{11} 出现的概率小于 2/5(否则无法满足 C_{01} 出现概率比 C_{10}, C_{11} 出现的概率大). 此时将 $00x$ 换成 $1x$，$1x$ 换成 $00x$，这样使得平均码长减少. 但是我们知道哈夫曼码是最优码，则这个改进与假设相矛盾，所以 $l_1 = 1$.

(b) 反证法. 假设分配给 p_1 的码字长度 $l_1 = 1$. 设 p_1 对应的码字以 0 开头，并以 C_x 表示以 x 开头的码字，由题意知 C_0 出现的概率小于 $1/3$.C_{10}, C_{11} 两个集合出现概率和大于 2/3，则其中必有一个集合概率大于 1/3，设为 C_{10}. 在此情况下，将 $0x$ 换为 $10x$，将 $10x$ 换为 $0x$，这样使得平均码长减少. 但是我们知道哈夫曼码是最优码，则这个改进与假设相矛盾，所以 $l_1 \geqslant 2$.

遇到这道题的时候，我通过构造码树进行反证. 我觉得构造码树会更直观一些. 下面分享我的做法.

(a) 反证法. 我们考虑码树构造过程，并且只考虑最后 3 次组合情况. 这样做是合理的，因为我们要讨论分配给 p_1 的码字长度，希望证明它不会超过 1，那我们的反证法就是说明其为 2 时矛盾，而这样在码树中码元组合的时候只和最后 3 次组合有关.

概率

p_1	p_1'	p_1'	1
p_2'	p_1	p_1''	
p_3'	p_2'		
p_4'			

由构造过程知

$$p_3' + p_4' > p_1 > \frac{2}{5}$$

则

$$p_2' = 1 - (p_3' + p_4') - p_1 < 1 - \frac{2}{5} - \frac{2}{5} = \frac{1}{5}$$

又 $p_2' > p_3' > p_4'$，则

$$p_3' + p_4' < 2 \times p_2' < \frac{2}{5}$$

则矛盾. 所以分配给 p_1 的码字长度小于 2，即为 1.

(b) 反证法. 类似上面的分析，构造码树讨论最后三次组合，寻找矛盾. 由于我们假设分配给 p_1 的码字长度为 1，则说明在概率合并的时候它除了最后一次以外，没有参与合并. 则说明

$$p_1 > p_2', \quad p_1 > (p_3' + p_4')$$

于是有

$$p_1 + p_2' + (p_3' + p_4') < 3 \times p_1 < 3 \times \frac{1}{3} = 1$$

则矛盾. 所以分配给 p_1 的码字长度必超过 1.

7.（12 分）考虑串联信道和并联信道的信道容量.

(a) 考虑 n 个完全相同的独立二元对称信道串联，如图所示.

$$X_0 \to \boxed{\text{BSC}} \to X_1 \to \cdots \to X_{n-1} \to \boxed{\text{BSC}} \to X_n$$

其中每个信道的原始误差概率为 p，证明该串联信道等价于一个二元对称信道，具有误差概率

$$p^* = \frac{1}{2}\left(1 - (1 - 2p)^n\right)$$

(b) 已知信道的转移概率矩阵为

$$P = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_N \end{bmatrix}$$

其中 $P_1, P_2, \ldots, P_N$ 是 N 个离散信道的信道矩阵. 令 $C_1, C_2, \ldots, C_N$ 表示 N 个离散信道的容量. 求该信道的容量.

解：

(a) 每个二元对称信道的转移概率矩阵为

$$A = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

则串联后的信道转移概率矩阵 $A^* = A^n$. 为了求解矩阵的 n 次幂，我们首先对矩阵

进行对角化 (在 8.1.6 节中详细介绍了矩阵对角化的方法). 这里直接给出结果

$$A = T^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1-2p \end{bmatrix} T$$

其中

$$T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

于是有

$$\begin{aligned} A_n &= A^n \\ &= T^{-1} \begin{bmatrix} 1 & 0 \\ 0 & (1-2p)^n \end{bmatrix} T \\ &= \begin{bmatrix} \frac{1}{2}\left(1+(1-2p)^n\right) & \frac{1}{2}\left(1-(1-2p)^n\right) \\ \frac{1}{2}\left(1-(1-2p)^n\right) & \frac{1}{2}\left(1+(1-2p)^n\right) \end{bmatrix} \end{aligned}$$

从转移概率矩阵中可得

$$p^* = \frac{1}{2}\left(1-(1-2p)^n\right)$$

(b) 由转移概率矩阵可知，该信道可以描述为 N 个离散信道 $P_1, P_2, \ldots, P_N$ 的并联，其中任意两个信道的输入输出字母表交集为 ϕ. 所以，问题可以描述为：求解并联信道的信道容量.

我们首先讨论两个信道并联的问题. 第一个信道的输入字母表为 $\mathcal{X}_1$，输出字母表为 $\mathcal{Y}_1$；第二个信道的输入字母表为 $\mathcal{X}_2$，输出字母表为 $\mathcal{Y}_2$；其中 $\mathcal{X}_1$ 和 $\mathcal{X}_2$ 交集为空，$\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 交集也为空 (就是两个“完全独立”的信道按照某一概率分布将输入组合在一起). 记第一个信道的信道容量为 C_1，第二个信道的信道容量为 C_2，需要求解并联后的信道容量 C.

两个信道组合方式可以描述为

$$X = \begin{cases} X_1, & \text{概率}\alpha \\ X_2, & \text{概率}(1-\alpha) \end{cases}$$

对于这种输入的随机变量按照一定概率分布组合在一起的问题，一般的处理技巧是引入示性变量. 记

$$\theta(X) = \begin{cases} 1, & X = X_1 \\ 2, & X = X_2 \end{cases}$$

引入示性变量的目的是通过构造互信息的分解，简化 $I(X;Y)$ 的表达，从而完成最大值求解.

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

由于给定了输入取值 ($X=X_1$ 或 $X=X_2$) 后输出 $Y=Y_1$ 还是 $Y=Y_2$ 是确定的，即 $\theta=g(Y)$. 所以构成马尔科夫链

$$X \rightarrow Y \rightarrow \theta$$

于是有 $I(X;\theta \mid Y)$=0. 则可以得到互信息的表达

$$\begin{aligned}
I(X;Y) &= I(X;\theta)+I(X;Y \mid \theta) \\
&= H(\theta)-H(\theta \mid X)+\alpha I\left(X_1;Y_1\right)+(1-\alpha) I\left(X_2;Y_2\right) \\
&= H(\alpha)+\alpha I\left(X_1;Y_1\right)+(1-\alpha) I\left(X_2;Y_2\right)
\end{aligned}$$

对上式进行求导可得

$$H'(\alpha)+C_1-C_2=0$$

其中

$$\begin{aligned}
H'(\alpha) &= -\log \alpha-\alpha \frac{1}{\alpha \ln 2}+\log(1-\alpha)+(1-\alpha) \frac{1}{(1-\alpha) \ln 2} \\
&= \log \frac{1-\alpha}{\alpha}
\end{aligned}$$

则可得

$$\log \frac{1-\alpha}{\alpha}=C_2-C_1$$

进而可得

$$\alpha=\frac{2^{C_1}}{2^{C_1}+2^{C_2}}$$

所以可得信道容量

$$\begin{aligned}
C &= H\left(\frac{2^{C_1}}{2^{C_1}+2^{C_2}}\right)+\frac{2^{C_1}}{2^{C_1}+2^{C_2}} C_1+\frac{2^{C_2}}{2^{C_1}+2^{C_2}} C_2 \\
&= -\frac{2^{C_1}}{2^{C_1}+2^{C_2}} \log \frac{2^{C_1}}{2^{C_1}+2^{C_2}}-\frac{2^{C_2}}{2^{C_1}+2^{C_2}} \log \frac{2^{C_2}}{2^{C_1}+2^{C_2}}+\frac{2^{C_1}}{2^{C_1}+2^{C_2}} C_1+\frac{2^{C_2}}{2^{C_1}+2^{C_2}} C_2 \\
&= \frac{1}{2^{C_1}+2^{C_2}}[-2^{C_1} \log 2^{C_1}+2^{C_1} \log(2^{C_1}+2^{C_2})-2^{C_2} \log 2^{C_2}+2^{C_2} \log(2^{C_1}+2^{C_2}) \\
&\quad +2^{C_1} C_1+2^{C_2} C_2] \\
&= \log\left(2^{C_1}+2^{C_2}\right)
\end{aligned}$$

达到信道容量时的输入概率分布为

$$X=\begin{cases} X_1 & \text{概率} \frac{2^{C_1}}{2^{C_1}+2^{C_2}} \\ X_2 & \text{概率} \frac{2^{C_2}}{2^{C_1}+2^{C_2}} \end{cases}$$

而其中 X_1 和 X_2 的概率分布为使得其对应信道达到信道容量的概率分布.

也可以做一个简单的变形得到一种容易记忆的形式

$$2^C = 2^{C_1} + 2^{C_2}$$

我们可以发现，这样的形式可以简单地进行推广，即对于若干个满足上述要求的“独立”信道并联，并联后的信道容量满足

$$2^C = \sum_{i=1}^{n} 2^{C_i}$$

所以本题，信道容量为

$$C = \log \sum_{i=1}^{N} 2^{C_i}$$

8.(6 分) 分别计算以下分布的微分熵

(a) 拉普拉斯分布 $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$

(b)$Y = X_1 + X_2$，其中 $(X_1, X_2) \sim \mathcal{N}\left(a, b, \sigma_1^2, \sigma_2^2, \rho\right)$.

解：

(a)

$$\begin{aligned}
h(f) &= -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|} \left[\ln\frac{1}{2} + \ln\lambda - \lambda|x|\right] dx \\
&= -\ln\frac{1}{2} - \ln\lambda + 1 \\
&= \ln\frac{2e}{\lambda} \text{ 奈特} \\
&= \log\frac{2e}{\lambda} \text{ 比特}
\end{aligned}$$

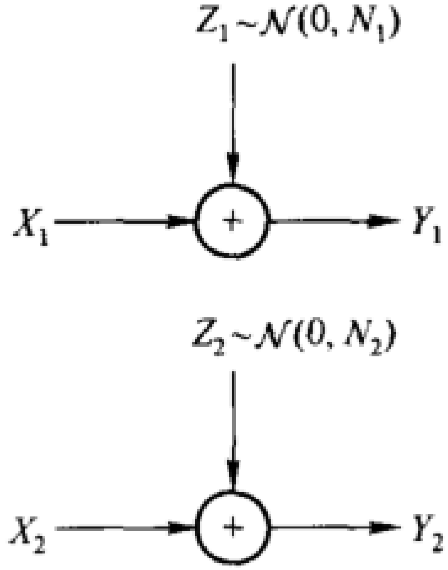
(b) 对于一般的变量和，可以利用卷积公式，或者引入辅助变量构造可逆变换，然后求解联合密度函数，再对辅助变量实数域积分得到我们要的“边缘概率密度函数”即可).这里由于是正态分布，可以直接使用结论

$$Y \sim \mathcal{N}\left(a + b, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right)$$

于是利用正态分布的微分熵可得

$$h(f) = \frac{1}{2}\log 2\pi e\left(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right) \text{ 比特}$$

9.(8 分) 考虑并联高斯信道



其中 $Z_1 \sim \mathcal{N}(0, N_1)$ 和 $Z_2 \sim \mathcal{N}(0, N_2)$ 为独立的高斯噪声. 选取固定的 β_1 和 β_2，满足 $\beta_1 N_1 > \beta_2 N_2$. 信道的功率分配约束为 $\beta_1 P_1 + \beta_2 P_2 \leqslant \beta$.

(a)β 取何值时信道停止单信道角色而开始起到双信道的作用?

(b)$\beta_1 = 1, \beta_2 = 2, N_1 = 3, N_2 = 2, \beta = 10$. 求解信道容量.

解析：

(a) 并联高斯信道模型，类似地可以得到

$$\beta = \beta_1 N_1 - \beta_2 N_2$$

(b) 求解出功率分配 $P_1 = 6 + 1 = 7$，$P_2 = 3$. 信道容量为

$$C = \frac{1}{2}\log(1 + \frac{7}{3}) + \frac{1}{2}\log(1 + \frac{3}{2}) = 1.53\text{比特/信道使用}$$

10.(8 分) 一个三维独立并联高斯信源 (X_1, X_2, X_3)，其中 X_1，X_2，X_3 均值都为零，方差分别是 2、8 和 4. 采用平方误差失真度量，$D = \sum_{i=1}^{3} D_i = \sum_{i=1}^{3} (x_i - \hat{x}_i)^2$，求该信源的信息率失真函数 $R(D)$.

解析：

直接由反注水法可得

- $0 < D < 2 \times 3 = 6$ 时：
  失真分配为平均分配 $(\frac{D}{3}, \frac{D}{3}, \frac{D}{3})$. 此时率失真函数为

$$R(D) = \sum_{i=1}^{3} \frac{1}{2}\log\frac{\sigma_i^2}{D_i} = \frac{1}{2}\log\frac{1728}{D^3}$$

- $6 \le D < 2+4\times 2 = 10$ 时：

  失真分配为 $(2, \frac{D}{2}, \frac{D}{2})$. 此时率失真函数为

$$R(D) = \sum_{i=1}^{3} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} = \frac{1}{2} \log \frac{64}{D^2} = \frac{8}{D}$$

- $10 \le D < 2+4+8 = 14$ 时：

  失真分配为 $(2, 4, D)$. 此时率失真函数为

$$R(D) = \sum_{i=1}^{3} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} = \frac{1}{2} \log \frac{8}{D}$$

- $D \ge 14$ 时：

  失真分配为 $(2, 4, D)$. 此时率失真函数为

$$R(D) = 0$$

综上所述，率失真函数为

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{1728}{D^3}, & 0 < D < 6, \\ \log\frac{8}{D}, & 6 \leqslant D < 10, \\ \frac{1}{2}\log\frac{8}{D}, & 10 \leqslant D < 14, \\ 0, & D > 14. \end{cases}$$

11.(6 分) 考虑在集合 $\{1, 2, \dots, m\}$ 上均匀分布的信源 X. 若失真度量为汉明失真，即

$$d(x, \hat{x}) = \begin{cases} 0, & \text{如果} x = \hat{x} \\ 1, & \text{如果} x \neq \hat{x} \end{cases}$$

求信源的率失真函数.

解：

首先我们对互信息进行分解

$$I(X; \hat{X}) = H(X) - H(X \mid \hat{X})$$

然后引入失真. 对于二元信源，采用汉明失真，引入失真的方法为费诺不等式

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

由题意知，$P_e = p(x \neq \hat{x}) = Ed(x, \hat{x}) = D$. 由费诺不等式有

$$H(X \mid \hat{X}) \leqslant H(D) + D\log(m-1)$$

接着求解条件熵的最大值即可得互信息最小值.

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X \mid \hat{X}) \\ &\geqslant \log m - H(D) - D\log(m-1) \end{aligned}$$

下面分析等号成立条件. 根据费诺不等式的等号成立条件可知

$$p(\hat{x} \mid x) = \begin{cases} 1-D, & \text{如果} \hat{x} = x \\ \dfrac{D}{m-1}, & \text{如果} \hat{x} \neq x \end{cases}$$

最后一步是确定临界值. 这一步比较简单 (但是容易忽略)，直接令前面求解的率失真函数为 0，得到对应的 D. 令率失真函数为 0，可得 $D = 1 - \dfrac{1}{m}$. 综上所述，率失真函数为

$$R(D) = \begin{cases} \log m - H(D) - D\log(m-1), & 0 \leqslant D \leqslant 1 - \dfrac{1}{m} \\ 0, & D > 1 - \dfrac{1}{m} \end{cases}$$