2.1 掷硬币。抛掷一枚均匀的硬币，直到第一次出现正面为止，设X表示所需的抛掷次数

(a)求熵 H(X)，单位为比特。下面的两个表达式可能会用到:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{(1-r)}, \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}$$

(b)假定随机变量X服从该分布。试找出一个“有效”的是否型问题序列，其问题形式如“X 包含于集合 S 吗，将 H(X)与确定 X 取值所需问题数的期望值进行比较。

解:

(a)

根据题意我们知道这是一个典型的几何分布：在伯努利试验中，记每次试验中事件 A 发生的概率为 p，试验进行到事件 A 出现时停止，此时所进行的试验次数为 X，其分布列为:

$P(X=k)=(1-p)^{k-1}p, k=1,2,\ldots$ 因此，我们有:

$$\begin{aligned}
&H(X)\\
&=-\sum_{n=1}^{\infty} p(1-p)^{n-1}\log(p(1-p)^{n-1})\\
&=-\left[\sum_{n=0}^{\infty} p(1-p)^{n-1}\log p+\sum_{n=0}^{\infty}(n-1)p(1-p)^{n-1}\log(1-p)\right]\\
&=\frac{-p\log p}{p}-\frac{(1-p)\log(1-p)}{p}\\
&=H(p)/p\ bits.
\end{aligned}$$

又 $p=1/2$， $H(X)=2\ bits$

（b）啥意思没读懂题目

2.2 函数的熵。设 X 是取有限个值的随机变量。如果

(a)$Y=2^X$ (b)$Y=cosX$

H(X)和 H(Y)的不等关系(或一般关系)是什么?

解：

事实上本题可直接利用下面的**习题 4**，即随机变量函数的熵，$Y=g(X)\Rightarrow H(Y)\le H(X)$ 当且仅当 X 和 g(X)一一对应时等式成立。但我们希望从另一个角度进行严格推导。

因为:

$H(X)=-\sum_{x} p(x)\log p(x)$ 因此，我们需要找到 $p(x)$ 与 $p(y)$ 之间的关系。

而根据题意 Y 都是关于 X 的函数，即 $y=f(x)$ ，而根据离散型随机变量的函数分布计算公

式： $p(y)=\sum_{x\in\{y=f(x)\}} p(x)$ ，**接下来的变形没想到，重新做。。。。**

2.3 最小熵。求$H(p_1,\dots,p_n)=H(\boldsymbol{p})$的最小值，其中$\boldsymbol{p}$的取值域为 n 维概率向量集合。请找出所有达到这个最小值时的$\boldsymbol{p}$。

根据熵的极值性我们知道，当取等概的时候，熵最大。而本题是想要我们找到最小值，根据 $H(P)=-\sum_i p_i \log p_i \geq 0$ 因此最小值如果能取到 0，那么本题就解决了。而我们知道当 $p_i=1$ 时，$p_j=0$, $\mathrm{i}\neq\mathrm{j}$，刚好取到。又因为 p_i 的取值范围时 n 维概率向量集合，因此

$p=(1,0,\dots,0),(0,1,\dots,0),\dots\dots,(0,0,\dots,1)$

2.4 随机变量的函数的熵，设X为离散型随机变量。请通过验证如下步骤证明X的函数的熵必小于或等于 X 的熵:

$$\begin{aligned}H(X,g(X)) &= H(X)+H(g(X)|X)\\ &= H(X)\end{aligned}$$

$$\begin{aligned}H(X,g(X)) &= H(g(X))+H(X|g(X))\\ &\geq H(X)\end{aligned}$$

因而有$H(g(X)) <= H(X)$。

解：

证明：因为

$$\begin{aligned}H(X,g(X)) &= H(X)+H(g(X)\,|\,X)\\ &= H(X)+\sum_x p(x)H(g(X)\,|\,X=x)\end{aligned}$$

当 X 确定之后 g(X)是被唯一确定的，因此$H(g(X)\,|\,X=x)=0$，从而，

$$\begin{aligned}H(X,g(X)) &= H(X)+H(g(X)\,|\,X)\\ &= H(X)+0 \qquad (1)\\ &= H(X)\end{aligned}$$

又因为

$H(X,g(X))=H(g(X))+H(X\,|\,g(X))$

又 $H(X\,|\,g(X))\geq 0$

$H(X,g(X))=H(g(X))+H(X\,|\,g(X)) \geq H(g(X))$ （2）

综（1）和（2）可得到$H(g(X))\leq H(X)$

2.5 零条件熵。证明:若$H(Y|X)=0$，则Y是X的函数(即对于满足$p(x)>0$的任意 x，仅存在一个可能取值y，使得$p(x,y)>0$)。

解：

证明：根据题意，我们首先想到用反证法来进行证明

假设对于任意满足 $p(x)>0$ ，存在 y_1,y_2 使得 $p(x,y_1)>0, p(x,y_2)>0$

又因为 $p(x)=\sum_y p(X=x,Y=y)\geq p(x,y_1)+p(x,y_2)>0$ ，因此

$$p(y_1\mid x)=\frac{p(x,y_1)}{p(x)}>0, p(y_2\mid x)=\frac{p(x,y_2)}{p(x)}>0$$

根据条件熵定义，我们有

$$H(Y\mid X)=-\sum_x p(x)\sum_y p(y\mid x)\log p(y\mid x)$$

$$\geq p(x)\left(-p(y_1\mid x)\log p(y_1\mid x)-p(y_2\mid x)\log p(y_2\mid x)\right)>0$$

这与题意矛盾，因此原命题得证。

2.6 条件互信息与无条件互信息。试给出联合随机变量 X,Y 和 Z 的例子，使得

(a) $I(X;Y|Z)<I(X;Y)$;

(b) $I(X;Y|Z)>I(X;Y)$;

解：

(a)

$$I(X;Y\mid Z)<I(X;Y)$$

这个不等式首先会让我们想到课本：Theorem 2.8.1 最后一个推论：If $X\to Y\to Z$ then $I(X;Y\mid Z)\leq I(X;Y)$ 当且仅当 X，Z 相互独立时等号成立。而我们恰恰不需要等号成立，因此我们需要找到 X 和 Z 相关的例子。因此我们假设 Y=Z=X,且 X 是服从伯努利分布 $B\sim(\frac{1}{2})$,则

$$I(X;Y)=H(X)-H(X\mid Y)=H(X)=1$$

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

因此有 $I(X;Y\mid Z)<I(X;Y)$

(b)

我们知道(a)中的不等式是本质上就是 Theorem 2.8.1 最后一个推论的特殊情况，条件是 X,Y,Z 构成马尔科夫链。那么我们很容易想到，当三者不构成马尔科夫链时，等式很可能不成立，即 $I(X;Y\mid Z)>I(X;Y)$，例如：设 X,Y 是相互独立的二元随机变量，$Z=X+Y$ 则

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

而

$$I(X;Y|Z)=H(X|Z)-H(X|Y,Z)$$
$$=H(X|Z)=P(Z=1)H(X|Z=1)=\frac{1}{2}bits$$

因此有：$I(X;Y|Z)>I(X;Y)$

2.7 硬币称重。假定有 n 枚硬币，可能有一枚或者没有假币。如果是假币，那么它的重量要么重于其他的硬币，要么轻于其他的硬币。用天平对硬币称重。

(a)若称重 k 次就能发现假币(如果存在)，且能正确判断出该假币是重于还是轻于其他硬币,试求硬币数 n 的上界。

(b)（较难)试给出对 12 枚硬币仅称$k=3$次就能发现假币的称重策略。

分析：根据熵的可加性，一个复合事件的不确定性可以通过多次试验逐步解除。如果每次实验所获的信息量都是可以获得的最大信息量，那么所需要的实验次数就可以最少。基于这个原理考虑这个问题，天平一次称重会有三种结果，且每一种结果的概率都是 1/3，那么进行一次称重所获的信息量是 x ，k 次称重所获得的最大信息量是 kx 。如果一共有 n 个硬币，且已知有一枚假币(不知道是否轻或重)，那么找到这枚假币所需的信息量是多少 ，就是我们需要求解的问题。

解：

(a)

方法一：天平一次称重会有三种结果（重，相等，轻），且每一种结果的概率都是 1/3 时，熵最大（等概的时候上最大）为：

$$H(X)=-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3}=\log_2 3$$

n 枚硬币，一共有 2n+1 中情况，每种情况发生概率相等，$\frac{1}{2n+1}$ 因此信息量为$\log_2(2n+1)$ ，

而每次使用天秤称，信息量就会少$\log_2 3$，因此需要称$k=\frac{\log_2(2n+1)}{\log_2 3}$

方法二：

本质上这其实是一个优化问题，根据我们之前的分析，可以建立一个优化模型

$$\min\ \ k$$
$$\text{s.t}\ k\log_2 3\geq\log_2(2n+1)$$
$$k\in N^{+},n\geq 3$$

我们只需要设计一个算法求解上述优化问题即可。

(b) 分析：12 枚硬币，需要称 3 次给出结果，我们需要设计三次称重方案：

这是一个很经典的数学问题，查了一些资料，还没整理好

2.8 有放回与无放回抽取。一个容器里面装有 r 个红球，w 个白球和 b 个黑球。若从容器中抽取 k 个球($k \geq 2$)，对有放回和无放回两种情形，哪种情形的熵更大?请回答并给予证明。

解：

当求有放回时，我们可以求得各个小球的概率，本质上时无条件熵，而有放回小球是条件熵，并且不是相互独立，根据，$H(X|Y) \leq H(X)$，条件作用使熵减小，因此，有放回条件熵更小。

2.9 度量，对于任意的 x 和 y，满足

$$\rho(x,y) \geq 0$$

$$\rho(x,y) = \rho(y,x)$$

当且仅当$x = y, \rho(x,y) = 0$

$$\rho(x,y) + \rho(y,z) \geq \rho(x,z)$$

则称函数$\rho(x,y)$为一个度量。

（a）证$\rho(X,Y) = H(X|Y) + H(Y|X)$满足上述第一条、第二条和第四条性质。如果存在从 X 到 Y 的一对一函数映射，我们说$X = Y$，$\rho(X,Y)$也满足也满足第三条性质，因而它是度量。

(b)验证$\rho(X,Y)$可表示为

$$\begin{aligned}\rho(X,Y) &= H(X) + H(Y) - 2I(X;Y) \\ &= H(X,Y) - I(X;Y) \\ &= 2H(X,Y) - H(X) - H(Y)\end{aligned}$$

解：

（a）

证明:

因为 $\rho(X,Y) = H(X|Y) + H(Y|X)$ 对于第一条因为条件熵大于等于 0，所以

$\rho(X,Y) \geq 0$

对于第二条：$\rho(X,Y) = H(X|Y) + H(Y|X) = H(Y|X) + H(X|Y) = \rho(Y,X)$得证。

对于第四条：

$\rho(X,Y) = H(X|Y) + H(Y|X)$

$\rho(Y,Z) = H(Y|Z) + H(Z|Y)$

$\rho(X,Z) = H(X|Z) + H(Z|X)$

$\Rightarrow \rho(X,Y) + \rho(Y,Z) = H(X|Y) + H(Y|X) + H(Y|Z) + H(Z|Y)$

因此我们只需证明

$H(X|Y) + H(Y|Z) \geq H(X|Z)$即可

因为

$H(X|Y) + H(Y|Z)$

对于第四条：根据第五题，我们知道当 Y 是 X 的函数时我们有 $H(X\mid Y)$ =0，因此当 x=y 时

$\rho(x,y)=0$

（b）

因为 $\rho(X,Y)=H(X\mid Y)+HH(X)-H(X\mid Y)(Y\mid X)$

又 $I(X;Y)=H(X)-H(X\mid Y)=I(Y;X)=H(Y)-H(Y\mid X)$

因此 $\rho(X,Y)=H(X)+H(Y)-2I(X;Y)$

由 $H(X,Y)=H(X)+H(Y\mid X)\Rightarrow I(X;Y)=H(X)+H(Y)-H(X,Y)$

因此有

$\rho(X,Y)=H(X)+H(Y)-2I(X;Y)$

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

=2H(X,Y)$-H(X)-H(Y)$

证毕。

2.10 不相交组合的熵。设离散型随机变量X_1,X_2的概率密度函数分别为$p_1(\cdot),p_2(\cdot)$字母表分别为$\mathcal{X}_1=\{1,2,\dots,m\},\mathcal{X}_2=\{m+1,\dots,n\}$。设

$$X=\begin{cases}X_1 & \textit{概率为}\alpha\\ X_2 & \textit{概率为}\, 1-\alpha\end{cases}$$

试求 H(X)关于$H(X_1),H(X_2)$和α表达式。

试对α进行最大化，证明$2^{H(X)}\le 2^{H(X_1)}+2^{H(X_2)}$ 利用$2^{H(X)}$为有效的字母表大小这个概念对此进行解释。

2.11 相关性的度圣。设X_1,X_2同分布，但不一定独立。设

$$\rho=1-\frac{(X_2|X_1)}{H(X_1)}$$

证明 $\rho=\frac{I(X_1;X_2)}{H(X_1)}$。

证明$0\le\rho\le 1$。

何时有$\rho=0$?

何时有$\rho=1$?

解：

（1）

$$\rho = 1 - \frac{H(X_2 | X_1)}{H(X_1)}$$
$$= \frac{H(X_1) - H(X_2 | X_1)}{H(X_1)}$$

又因为 X_1 和 X_2 同分布，因此有 $H(X_1) = H(X_2)$ ，因此

$$\rho = 1 - \frac{H(X_2 | X_1)}{H(X_1)} = \frac{H(X_1) - H(X_2 | X_1)}{H(X_1)}$$
$$= \frac{H(X_2) - H(X_2 | X_1)}{H(X_1)}$$
$$= \frac{I(X_1; X_2)}{H(X_1)}$$

(2)

因为 $I(X_1; X_2) = H(X_2) - H(X_2 | X_1) \geq 0$

因此有

$$0 \leq H(X_2 | X_1) \leq H(X_2) = H(X_1)$$
$$\Rightarrow 0 \leq \frac{H(X_2 | X_1)}{H(X_1)} \leq 1$$

$$\Rightarrow 0 \leq \rho = 1 - \frac{H(X_2 | X_1)}{H(X_1)} \leq 1$$

(3)

$$\rho = 0 \Rightarrow \rho = \frac{I(X_2; X_1)}{H(X_1)} \Rightarrow I(X_2; X_1) = 0$$

根据定理 2.6.3 推论（互信息的非负性）对任意两个随机变量 X 和 Y 有：$I(X_2; X_1) \geq 0$ 当且仅当 X 和 Y 相互独立时等号成立。

(4)

$\rho = 1 \Rightarrow \rho = \dfrac{H(X_2) - H(X_2 | X_1)}{H(X_1)} \Rightarrow H(X_2 | X_1) = 0$ 根据题⑤**当且仅当** X_1 是 X_2 的函数时，也就是一一映射的函数关系，反之也可以。

2.12 联合熵的例子。设$\rho(x, y)$由右表给出，

试计算：

X \ Y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

(a)$H(X)$，$H(Y)$

(b)$H(X|Y), H(Y|X)$

(c)$H(X, Y)$

(d)$H(Y)-H(Y|X)$

(e)$I(X;Y)$

(f)画出(a)~(e)中所有量的文氏图。

解：

（a）根据边际概率公式我们有

$$p(x=0)=\sum_{y\in Y}p(0,y)=\frac{2}{3}$$

$$p(x=1)=\sum_{y\in Y}p(1,y)=\frac{1}{3}$$

$$\Rightarrow H(X)=-\frac{2}{3}\log\frac{2}{3}-\frac{1}{3}\log\frac{1}{3}=H(Y)$$

（b）

$$p(y=0)=\sum_{x\in X}p(x,0)=\frac{1}{3}$$

$$p(y=1)=\sum_{x\in X}p(x,1)=\frac{2}{3}$$

$$\begin{aligned}\Rightarrow H(X\mid Y)&=P(y=0)H(X\mid Y=0)+P(y=1)H(X\mid Y=1)\\&=P(y=0)\sum_{x\in X}p(x\mid y)\log p(x\mid y)+P(y=1)\sum_{x\in X}p(x\mid y)\log p(x\mid y)\end{aligned}$$

又因为

$$p(x\mid y)=\frac{p(x,y)}{p(y)}$$

$$\Rightarrow p(x=0\mid y=0)=\frac{p(x=0,y=0)}{p(y=0)}=\frac{\frac{1}{3}}{\frac{1}{3}}=1$$

$$\Rightarrow p(x=1\mid y=0)=\frac{p(x=1,y=0)}{p(y=0)}=\frac{0}{\frac{1}{3}}=0$$

$$\Rightarrow p(x=0\mid y=1)=\frac{p(x=0,y=1)}{p(y=1)}=\frac{\frac{1}{3}}{\frac{2}{3}}=\frac{1}{2}$$

$$\Rightarrow p(x=1\mid y=1)=\frac{p(x=1,y=1)}{p(y=1)}=\frac{\frac{1}{3}}{\frac{2}{3}}=\frac{1}{2}$$

综上$H(X\mid Y)=\frac{2}{3}$　（$H(X\mid Y)$同求）

（c）

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \\
&= p(x=0,y=0)\log p(x=0,y=0) + p(x=0,y=1)\log p(x=0,y=1) \\
&+ p(x=1,y=0)\log p(x=1,y=0) + p(x=1,y=1)\log p(x=1,y=1) \\
&= 3\times\frac{1}{3}\log 3 = \log 3
\end{aligned}
$$

(d) 根据（a）和（b）即可得 $H(Y)-H(Y\mid X)$ ，这里不再赘述。

(e) $I(X;Y)=H(X)-H(X\mid Y)=H(Y)-H(Y\mid X)$ (同 e 中的答案)

(f) 课本上有，图就省略了。

2.13 不等式。证明对任意的$x>0, \ln x \geq 1-\frac{1}{x}$。

解：

此题有多种解法，在这里利用相对简单的一种即求导

令

$$
f(x)=\ln x+\frac{1}{x}-1
$$

$$
\Rightarrow f'(x)=\frac{x-1}{x^2}
$$

因此在 $x\in(0,1)$ $f(x)$ 单调递减 $x\in[1,\infty]$ $f(x)$ 单调递增，因此 $f(x=1)=0$ 是最最小值

因此

$$
f(x)\geq 0
$$

$$
\Rightarrow \ln x \geq 1-\frac{1}{x}
$$

2.14 和的熵。设随机变量 X，Y 的取值分别为$x_1,x_2,\dots\ ,x_r$和$y_1,y_2,\dots,y_s$ 设$Z=X+Y$。

(a)证明$H(Z|X)$=$H(Y|X)$，并讨论如果 X，Y 独立；则$H(Y)\leq H(Z)$及$H(X)\leq H(Z)$

由此说明独立随机变量的和增加不确定度。

(b)给出一个(必须是相关)随机变量例子，使得$H(X)>H(Z)$且$H(Y)>H(Z)$。

在什么条件下 $H(Z)=H(X)+H(Y)$?

解：

(a)

分析：这种我们只能根据定义来证明，要么证：左=右或者右=左

解：

$$
\begin{aligned}
H(Z\mid X) &= -\sum_{x\in X} p(x)H(Z\mid X=x) \\
&\Rightarrow -\sum_{x\in X} p(x)\sum_{z\in Z} p(z=x+y\mid x)\log(z=x+y\mid x)
\end{aligned}
$$

由概率论基本知识我们知道当 $Z = X + Y$，有 $p(z = x + y \mid x) = p(y = z - x \mid x)$

（b）

根据题目已知信息，我们可以获得 $H(X) \geq 0, \mathrm{H(Y)} \geq 0$ 要严格保证 $H(X) > H(Z), \mathrm{H(Y)} > H(Z)$ 最稳妥的就是 $H(Z) = 0$，即保证 Z 是一个已知的确定事件。因此我们尝试构造

随机变量 $X = -Y$ 且

$$P(X = -1) = \frac{1}{2} \Rightarrow P(Y = 1) = \frac{1}{2}$$

$$P(X = 1) = \frac{1}{2} \Rightarrow P(Y = -1) = \frac{1}{2}$$

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1 = H(Y)$$

因此 Z 必然等于 0.因此 $H(Y) = 0$

（c）

2.15 数据处理。设 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 依序构成马尔可夫链，即设

$$p(x_1, x_2, \dots, x_n) = p(x_1)\rho(x_2|x_1)\cdots p(x_n|x_{n-1})$$

试将 $I(X_1; X_2, \dots, X_n)$ 简化到最简单形式。

解：根据互信的链式法则我们有

$$I(X_1, X_2, \dots, X_n; Y) = I(X_1; Y) + I(X_2; Y \mid X_1) + I(X_3; Y \mid X_1, X_2), \dots\dots I(X_n; Y \mid X_1, X_2, \dots, X_{n-1})$$
$$\Rightarrow I(X_1; X_2, \dots, X_n) = I(X_1; X_2) + I(X_1; X_2 \mid X_3) + I(X_1; X_4 \mid X_2, X_3) \quad (3)$$

又因为 $X_1 \to X_2 \to X_3 \to \dots\dots \to X_n$ 依序构成马尔科夫链，根据马尔可夫链得性质：

即 t+1 步的随机变量在给定第 t 步随机变量后与其余的随机变量条件独立（conditionally independent）即 $P(X_{t+1} \mid X_t, X_{t-1}, \dots, X_1) = P(X_{t+1} \mid X_t)$

因此（3）中条件互信息全部为零，因此 $I(X_1; X_2, \dots, X_n) = I(X_1; X_2)$

备注：这里补充一下，根据定理 2.6.3（信息不等式的推论）：1：对任意两个随机变量 X 和 Y $I(X;Y) \geq 0$ 当且仅当 X 和 Y 相互独立等号成立。2：$I(X;Y \mid Z) \geq 0$ 当且仅当对给定随机变量 Z, X 和 Y 是条件独立得，等号成立。具体证明参加课本，利用相对熵直接证明即可。

2.16 瓶颈模型。假定(非平稳)马尔可夫链起始于 n 个状态中的一个，然后第二步受到限制，转移到 k 个状态之一($k < n$)，第三步又放宽转移到 m 个状态中的一个(m>k)于是有 $X_1 \to X_2 \to X_3$，即对任意的 $x_1 \in \{1,2,\dots,n\}, x_2 \in \{1,2,\dots,k\}, x_3 \in \{1,2,\dots,m\}$ 有

$p(x_1,x_2,x_3\ )=p(x_1)\rho(x_2|x_1)p(x_3|x_2)$。

(a)试通过证明$I(X_1;X_3)\le\log k$,说明X_1与X_3 的相关程度受瓶颈作用的限制情况。

(b)当$k=1$时$I(X_1;X_3)$并且得出结论:通过该瓶颈作用后X_1与X_3不再具有相关性。

解:

(a)

根据定理 2.8.1（数据处理不等式）若 $X\to Y\to Z$ ，则有 $I(X;Z)\le I(X;Y)$

因此我们可以得到

$$\begin{aligned}I(X_1;X_3)\le I(X_1;X_2)&=H(X_2)-H(X_2\mid X_1)\\&\le H(X_2)\end{aligned}$$

因此接下来我们只需要证明

$H(X_2)\le\log k$ 即可，本质上是证明 $H(X_2)=H(x_1,x_2,\ldots,x_k)\le\log k$

等价证明：$H(\mathrm{P}_1,\mathrm{P}_2,\ldots,\mathrm{P}_q)\le\log\mathrm{q}$ **当** $\mathrm{P}_1=\mathrm{P}_2\ldots=\mathrm{P}_q=\frac{1}{q}$ 时，信源具有最大熵。

先引入引理：当 $x>0$ 则：$\ln x\le x-1$ 即 $\log_2 x\le(x-1)\log e$ 当且仅当 x=1 时成立

$$\begin{aligned}H(\mathrm{P}_1,\mathrm{P}_2,\ldots,\mathrm{P}_q)-\log\mathrm{q}&=-\sum_{i=1}^{q}p_i\log p_i-\sum_{i=1}^{q}p_i\log q\\&=-\sum_{i=1}^{q}p_i\left(\log p_iq\right)=\sum_{i=1}^{q}p_i\frac{1}{\log p_iq}\end{aligned}$$

令 $x=\dfrac{1}{\log p_iq}$ 则

$$\begin{aligned}H(\mathrm{P}_1,\mathrm{P}_2,\ldots,\mathrm{P}_q)-\log\mathrm{q}&=-\sum_{i=1}^{q}p_i\log p_i-\sum_{i=1}^{q}p_i\log q\\&=-\sum_{i=1}^{q}p_i\left(\log p_iq\right)=\sum_{i=1}^{q}p_i\frac{1}{\log p_iq}\\&\le\sum_{i=1}^{q}p_i\left(\frac{1}{p_iq}-1\right)\log e=0\end{aligned}$$

因此

$$\begin{aligned}I(X_1;X_3)\le I(X_1;X_2)&=H(X_2)-H(X_2\mid X_1)\\&\le H(X_2)\le\log k\end{aligned}$$

综上 X_1 与 X_3 的相关度受瓶颈作用的影响。

(b)

因为当 $k=0$ 时

$I(X_1;X_3)\le\log k=0$ 又因为 $I(X_1;X_3)\ge0$

因此 $I(X_1;X_3)=0$ 所以 X_1 与 X_3 是相互独立的。

2.17 纯随机性 与倾向性硬币。设$X_1,X_2,\dots,X_n$表示独立地抛掷一枚倾向性硬币所产生的可能结果的随机变量。于是,$Pr\{X_i=1\}=p$，$Pr\{X_i=0\}=1-p$，其中p未知。要从$X_1,X_2,\dots,X_n$中获得均匀硬币抛掷的序列$Z_1,Z_2,\dots,Z_K$,为此，设$f:\mathcal{X}^n\to\{0,1\}^*$（其中$\{0,1\}^*=\{\Lambda,0,1,00,01,\dots\}$为所有有限长度的二元序列集合)表示映射$f(X_1,X_2,\dots,X_n)=(Z_1,Z_2,\dots,Z_K,)$,其中$Z_i\sim$ Bermoullii($\frac{1}{2}$),而 K 的取值可能依赖于$X_1,X_2,\dots,X_n$。为了让$Z_1,Z_2,\dots$成为抛掷均匀硬币所产生的随机序列,从倾向性硬币抛掷到均匀硬币抛掷的映射f必须具有特定的性质，即在给定长度k时，所有2^k个序列$(Z_1,Z_2,\dots,Z_k)$具有相同的概率(可能为 0),其中 k=1,2,。例如,n=2 时，映射$f(01)=0,f(10)=1,f(00)=f(11)=\Lambda$(空串),则有$Pr\{Z_1=1|K=1\}=Pr\{Z_1|K=1\}=\frac{1}{2}$=寸。请给出下列不等式成立的理由:

$$
\begin{aligned}
nH(p) &\overset{(a)}{=} H(X_1,\dots,X_n)\\
&\overset{(b)}{\geq} H(Z_1,Z_2,\dots,Z_K,K)\\
&\overset{(c)}{=} H(K)+H(Z_1,Z_2,\dots,Z_K|K)\\
&\overset{(d)}{=} H(K)+E(K)\\
&\overset{(e)}{\geq} EK
\end{aligned}
$$

因而在平均意义上,从$(X,\dots,X_n)$中得到的均匀硬币抛掷次数不会超过$nH(p)$。举出长度为4 的序列上的恰当的映射f。

解：

（a）

因为 $X_1,X_2,\dots X_n$ 相互独立，根据定理 2.6.6（*Independence bound on entropy*）:

Let $X_1,X_2,\dots X_n$ *be drawn according to* $p(x_1,x_2,\dots,x_n)$ *Then*

$H(X_1,X_2,\dots X_n)\leq\sum_{i=1}^{n}H(X_i)$ *with equality if and only if the Xi are independent.*

条件满足，又$\Pr(X_i=1)=p$ *所以*$nH(p)=H(X_1,X_2,\dots X_n)$

（b）

根据题意我们知道$Z_1,Z_2,\dots Z_n$是 $X_1,X_2,\dots X_n$ 的函数，根据之前题目 2.4，我们有随机变量的函数熵，即：$H(X_1,X_2,\dots X_n)\geq H(Z_1,Z_2,\dots Z_n)$

有问题这里：K 的取值依赖$(X_1,X_2,\dots X_n)$**?**

(c)
(d)
(e)

2.18 世界职业棒球锦标赛。世界职业棒球锦标赛为 7 场系列赛制，只要其中一队赢得 4 场，比赛就结束。设随机变量 X 代表在棒球锦标赛中，A 队和 B 队较量的结果。例如，X 的取值可能为 AAAA, BABABAB, BBBAAAA。设 Y 代表比赛的场数,取值范围为 4~7。假定 A 队和 B 队是同等水平的，且每场比赛相互独立。试计算$H(X)$, $H(Y)$,$H(Y|X)$及$H(X|Y)$。

解：

要求$H(X)$ 和$H(Y)$我们必须要求得$p(x_i)$ 和$p(y_i)$

首先对于 X 的可能取值情况我们有
4 场比赛 X 的可能性：

$$P(AAAA)=\left(\frac{1}{2}\right)^4$$

$$P(BBBB)=\left(\frac{1}{2}\right)^4$$

5 场比赛 X 的可能性：8=$2C_4^1$ 种可能

$$P(ABBBB)=\left(\frac{1}{2}\right)^5$$

$$P(BABBB)=\left(\frac{1}{2}\right)^5$$

$$P(BABBB)=\left(\frac{1}{2}\right)^5$$

$$\vdots$$

6 场比赛 X 的可能性：$2C_5^2=20$ 种，每种可能性为$\left(\frac{1}{2}\right)^6$

7 场比赛 X 的可能性：$2C_6^3=40$ 种，每种可能性为$\left(\frac{1}{2}\right)^7$

对于 Y 来说有以下可能性

$$P(Y=4)=2\frac{1}{2^4}=\frac{1}{8}$$

$$P(Y=5)=8\frac{1}{2^5}=\frac{1}{4}$$

$$P(Y=6)=20\frac{1}{2^6}=\frac{5}{16}$$

$$P(Y=7)=40\frac{1}{2^7}=\frac{5}{16}$$

因此 $H(X) = -2\frac{1}{2^4}\log\frac{1}{2^4} - 8\frac{1}{2^5}\log\frac{1}{2^5} - 20\frac{1}{2^6}\log\frac{1}{2^6} - 40\frac{1}{2^7}\log\frac{1}{2^7}$

$$H(Y) = -\frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - 2\frac{5}{16}\log\frac{5}{16}$$

根据上述分析我们知道 Y 本质上是关于 X 的函数，因此 $H(Y \mid X) = 0$

因为

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$
$$\Rightarrow H(X \mid Y) = H(X) + H(Y \mid X) - H(Y) = H(X) - H(Y)$$

将（1）中结果带入即可。

2.19 无穷熵。此题说明离散型随机变量的熵可能是无穷的。设$A = \sum_{n=2}^{\infty}(nlog^2 n)^{-1}$。(考虑到$(xlog^2 x)^{-1}$的积分为 A 的一个上界，容易证明 A 是有限的。)证明:设 X 是由$Pr(X = n) = (Anlog^2 n)^{-1}$定义的整数值随机变量，其中$n = 2,3,..$,则$H(X) = +\infty$。

解:

因为 $H(X) = -\sum_{n=2}^{\infty} p(n)\log p(n)$

根据题意有 $p(n) = \left(An\log^2 n\right)^{-1}$ 带入上式我们有:

$$\begin{aligned}
H(X) &= -\sum_{n=2}^{\infty} p(n)\log p(n) \\
&= \sum_{n=2}^{\infty} \frac{\log\left(An\log^2 n\right)}{\left(An\log^2 n\right)} \\
&= \sum_{n=2}^{\infty} \frac{\log A + \log n + 2\log(\log n)}{\left(An\log^2 n\right)} \\
&= \sum_{n=2}^{\infty} \frac{\log A + \log n + 2\log(\log n)}{\left(An\log^2 n\right)} \\
&= \sum_{n=2}^{\infty} \frac{\log A)}{\left(An\log^2 n\right)} + \sum_{n=2}^{\infty} \frac{\log n}{\left(An\log^2 n\right)} + \sum_{n=2}^{\infty} \frac{2\log(\log n)}{\left(An\log^2 n\right)}
\end{aligned}$$

根据题意我们很容易判别出来第一项和第三项都是有界得，因此我们的目标主要是处理中间项 $\sum_{n=2}^{\infty} \frac{\log n}{\left(An\log^2 n\right)} = \sum_{n=2}^{\infty} \frac{1}{\left(An\log n\right)}$

针对上式我们可以将离散形式等价转化成连续形式，即:

A 有界，我们设为 $A \le M$ ，因此有

$$\sum_{n=2}^{\infty} \frac{1}{\left(An\log n\right)} \ge \frac{1}{M}\int_{2}^{\infty} \frac{1}{x\log x}dx = \frac{1}{M}\log\log x\,|_2^{\infty} = \infty$$

因此原命题得证。

2.20 游程编码。设$X_1, X_2, \dots, X_n$(可能相关)均为二元随机变量。假定某人对此序列(按先后产生的次序)计算出游程$R = (R_1, R_2, \dots)$。例如，序列 X= 0001100100 产生游程为 R=(3,2,2,1,2)。请你比较$H(X_1, X_2, \dots, X_n), H(R)$及$H(X_n, R)$,给出所有等式和不等式关系以及差别的范围。

解:

(a)

根据题意我们可判断出 R 是关于 X 的函数，因此根据本节题目 4 中的结论;随机变量函数的熵小于等于随机变量的熵，因此$H(R) \le H(X)$

(b)

因为 $H(X_1, X_2, .., X_n) = \sum_{i=1}^{n} H\left(X_i \mid X_{i-1}, ..., X_1\right)$

而$H(X_n, R) = H(R) + H(X_n \mid R)$

又因为$H(X_n \mid R) \le H(X_n)$

$$\begin{aligned} H(X_n, R) &= H(R) + H(X_n \mid R) \\ &\le H(R) + H(X_n) \end{aligned}$$

为何没找$H(X_1, X_2, .., X_n)$与$H(X_n, R)$关系?

2.21 概率的马尔可夫不等式。设 p(x)为概率密度函数。证明对任意的$d \ge 0$,有

$$Pr\{p(X) \le d\} log \frac{1}{d} \le H(X)$$

解:

证明:

根据马尔可夫不等式

$$\Pr\{p(x) \le d\} \log \frac{1}{d} = \sum_{x \in \{p(x) \le d\}} p(x) \log \frac{1}{d} \le \sum_{x \in \{p(x) \le d\}} p(x) \log \frac{1}{p(x)} \le \sum_{x} p(x) \log \frac{1}{p(x)} = H(X)$$

综上原命题得证。

2.22 思路的逻辑顺序。在实际中,常常会由于某种需要而有序地论述某些思路，然后,若有必要就会对这些思路作进-一步的推广。请重新给如下所述思路排列顺序,要求是强的排在前面，蕴含的紧随其后。

(a) $I(X_1, X_2, \dots, X_n; Y)$的链式法则，$D(p(x_1, \dots, x_n) || q(x_1, x_2, \dots, x_n))$的链式法则，以及$H(X_1, X_2, .., X_n)$的链式法则。

(b) $D(f||g) \ge 0$,Jensen 不等式$I(X;Y) \ge 0$。

解:

(a)

由 Theorem 2.5.2 (Chain rule for information)

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

根据其证明过程:

$$I(X_1,X_2,...,X_n;Y)=H(X_1,X_2,...,X_n)-H(X_1,X_2,...,X_n\mid Y)$$

$$=\sum_{i=1}^{n}H(X_i\mid X_{i-1},...,X_1)-\sum_{i=1}^{n}(X_i\mid X_{i-1},...,X_1,Y)$$

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

我们可以看出 H 的链式法则强于 I。

另又 $I(X_1,X_2,...,X_n;Y)=D(p(x_1,x_2,...,x_n)\,\|\,q(x_1,x_2,...,x_n))$

因此有 D 的链式法则强于 I。

综上，逻辑顺序可分为两组：H,I； D；I

(b)

根据 $I(X;Y)=D(p(x,y)\,\|\,p(x)p(y))\geq 0$

因此 D 强于 I。

2.23 条件互信息。考虑 n 个二元随机变量X_1 ,X_2,..,X_n组成的序列。如果含偶数个 1 的每个序列的概率为$2^{-(n-1)}$，含奇数个 1 的每个序列的概率为 0,试计算以下的互信息

$$I(X_1;X2),I(X_2;X_3|X_1),..,I(X_n-1;X_n\ |X_1,\dots,X_{n-2})$$

解:

没读懂题目

2.24 平均熵。 设$H(p)=\ -\ plog_2p-(1\ -\ p)\log_2(1-\ p)$为二元熵函数。

(a)利用$\log_2 3\ \approx 1.584$,计算 H(1/4)的值。(提示:可以考虑具有 4 种等可能结果的试验，其中某个结果比其他的更有趣。)

(b)当概率 p 的值在 0≤p≤1 范围内均匀选取，试计算平均熵$H(p)$。

(c) (选做)试计算平均熵$H(p_1,p_2,p_3)$,其中(p_1,p_2,p_3)为均匀分布的概率向量。推广到 n 维情形。

解:

(a)

$$H(p)=-p\log_2 p-(1-p)\log_2(1-p)=\frac{1}{4}\log_2 4-\frac{3}{4}\log_2\frac{3}{4}=\frac{1}{2}-\frac{3}{4}(\log_2 3-\log_2 4)=0.811bits$$

(b) 平均熵，本质就是期望，因此

$$-\int_0^1 \left(p\log_2 p+(1-p)\log_2(1-p) \right) dp = -\frac{1}{\ln 2} 2\int_0^1 \left(p\ln p \right) dp = \frac{1}{2\ln 2}$$

(c)

2.25 文氏图。事实上，不存在度量三个随机变量所共有的互信息概念。在这里，我们尝试给出一种定义:根据文氏图,三个随机变量 X, Y 和 Z 的公共部分的互信息可定义为

$$I(X;\ Y;Z) =\ \ I(X;\ \ Y) -\ \ I(X;Y|Z)$$

尽管上述定义并不对称，其实这个量对于 X，Y，Z 是对称的，遗憾的是，$I(X;Y;Z)$并不一定非负，试举例 X，Y，Z，使得$I(X;Y;Z) < 0$，并证明以下两个恒等式：

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

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

找出 X,Y 和 Z 之间的关系使得$I(X;Y;Z)<0$ 第一个恒等式可类似的由熵和互信息的文氏图得到理解。第二个恒等式由第一个容易得到。

解：

(a)

根据$I(X;Y;Z)$的定义

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

原命题得证

(b)

根据**(a)**我们有

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

原命题得证

2.26 相对熵的非负性的另一个证明。为突出结论$D(p||q) \geq 0$的基本性，我们在各处另一个证明。

(a)证明对任意的$0 < x < \infty$，$\ln x \ \ \leq x - 1$

(b)判定下列步骤

$$
\begin{aligned}
-D(p||q) &= \sum_x p(x)\ln\left(\frac{q(x)}{p(x)}\right) \\
&\le \sum_x p(x)\left(\frac{q(x)}{p(x)}-1\right) \\
&\le 0
\end{aligned}
$$

(c)等号成立的条件是什么?

解:

(a)

非常简单，我们只需要构造 $f(x)=\ln x-x+1$

因为 $f'(x)=\frac{1}{x}-1$ 令 $f'(x)=0\Rightarrow x=1$，因此我们可以得到:

当 $x\in(0,1)\Rightarrow f'(x)>0$

当 $x\in[1,\infty)\Rightarrow f'(x)\le 0$

因此当 x=1 时 f(x)取得最大值，又 f(1)=0,因此 $f(x)\le 0\Rightarrow \ln x\le x-1$

证毕

(b)

(c) 本质就是 $\ln x\le x-1$ 什么时候等式成立，显然 x=1 得时候，对于本题，即

$\frac{q(x)}{p(x)}=1\Rightarrow q(x)=p(x)$ 时成立

2.27 熵的组合法则。设 $p=(p_1,p_2,\dots,p_m)$ 为 m 个元素上的概率分布(即 $p_i\ge 0$,且 $\sum_{i=1}^m p_i=1$)。定义 m-1 个元素上的新分布 $\boldsymbol{q}$ 为 $q_1=p_1, q_2=p_2,\dots,q_{m-2}=p_{m-2}$ 以及 $q_{m-1}=p_{m-1}+p_m$ (即分布 **q** 与 **p** 在集合 $\{1,2,\dots,m-2\}$ 上是相同的，**q** 中最后一个元素的概率为 **p** 中最后两个元素的概率之和)。证明

$$
H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1}+p_m)H\left(\frac{p_{m-1}}{p_{m-1}+p_m},\frac{p_m}{p_{m-1}+p_m}\right)
$$

解:

根据熵的定义我们有

$$
\begin{aligned}
H(p) &= -\sum_{i=1}^{m} p_i\log p_i = -\sum_{i=1}^{m-2} p_i\log p_i - p_{m-1}\log p_{m-1} - p_m\log p_m \\
&= -\sum_{i=1}^{m-2} p_i\log p_i - p_{m-1}\log\left(\frac{p_{m-1}}{p_{m-1}+p_m}(p_{m-1}+p_m)\right) - p_m\log\left(\frac{p_m}{p_{m-1}+p_m}(p_{m-1}+p_m)\right) \\
&= -\sum_{i=1}^{m-2} p_i\log p_i - p_{m-1}\log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - p_m\log\left(\frac{p_m}{p_{m-1}+p_m}\right) - (p_{m-1}+p_m)\log(p_{m-1}+p_m) \\
&= -\sum_{i=1}^{m-2} p_i\log p_i - p_{m-1}\log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - p_m\log\left(\frac{p_m}{p_{m-1}+p_m}\right) - (q_{m-1})\log(p_{m-1})
\end{aligned}
$$

$$= H(q) - (p_{m-1} + p_m)\left(\frac{p_{m-1}}{p_{m-1}+p_m}\log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) + \frac{p_m}{p_{m-1}+p_m}\log\left(\frac{p_m}{p_{m-1}+p_m}\right)\right)$$

$$= H(q) - (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right)$$

综上原命题得证.

2.28 混合使熵增加。 证明概率分布$(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$的熵小于概率分布

$(p_1, \ldots, (p_i + p_j)/2, \ldots, (p_i + p_j)/2, \ldots, p_m)$的熵。进一步证明更一般的结论:使概率分布更均匀的变换都使熵增加。

解:

根据熵得定义:

首先我们令

$$P = (p_1, \ldots p_i, \ldots p_j, \ldots, p_m)$$

$$Q = (p_1, \ldots \frac{p_i + p_j}{2}, \ldots \frac{p_i + p_j}{2}, \ldots, p_m)$$

我们有:

$$H(P) = -\sum_{i=1}^{m} p_i \log p_i$$

$$H(P) - H(Q) = -p_i \log p_i - p_j \log p_j + (p_j + p_i)\log\frac{p_j + p_i}{2} < 0$$

因此$H(P) \le H(Q)$

更一般的结论: **当** $\mathrm{P}_1 = \mathrm{P}_2 \ldots = \mathrm{P}_q = {}^1\!/_q$ 时，信源具有最大熵。

证明: 本质上是熵的极值性

熵的极值性

$H(\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_q) \le \log \mathrm{q}$ **当** $\mathrm{P}_1 = \mathrm{P}_2 \ldots = \mathrm{P}_q = {}^1\!/_q$ 时，信源具有最大熵。

证明:

先引入引理: 当 $x > 0$ 则: $\ln x \le x - 1$ 即$\log_2 x \le (x-1)\log e$ 当且仅当 x=1 时成立

$$H(\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_q) - \log \mathrm{q} = -\sum_{i=1}^{q} p_i \log p_i - \sum_{i=1}^{q} p_i \log q$$

$$= -\sum_{i=1}^{q} p_i (\log p_i q) = \sum_{i=1}^{q} p_i \frac{1}{\log p_i q}$$

令 $x=\dfrac{1}{\log p_i q}$ 则

$$
\begin{aligned}
H(\mathrm{P}_1,\mathrm{P}_2,\ldots,\mathrm{P}_q)-\mathrm{logq}&=-\sum_{i=1}^{q}p_i\log p_i-\sum_{i=1}^{q}p_i\log q\\
&=-\sum_{i=1}^{q}p_i\left(\log p_i q\right)=\sum_{i=1}^{q}p_i\frac{1}{\log p_i q}\\
&\le\sum_{i=1}^{q}p_i\left(\frac{1}{p_i q}-1\right)\log e=0
\end{aligned}
$$

这就意味着任何一个信源 q 元熵减去 logq 是小于等于 0 的。当且仅当 $p_i=\dfrac{1}{q}$ 时成立。

2.29 不等式。设 X、Y 和 Z 为联合随机变量。证明下面的不等式，并给出等号成立的条件。

(a) $H(X,Y|Z)\ge H(X|Z)$。

(b) $I(X,Y;Z)\ge I(X;Z)$。

(c) $H(X,Y,Z)-H(X,Y)\le H(X,Z)-H(X)$。

(d) $I(X;Z|Y)\ge I(Z;Y|X)-\ I(Z;Y)+\ I(X;Z)$。

解：

(a)

$$H\left(X,Y\mid Z\right)\ge H(X\mid Z)$$

根据 **Theorem 2.2.1** (*Chain rule*) **Corollary，**

$$H\left(X,Y\mid Z\right)=H\left(X\mid Z\right)+H(Y\mid X,Z)\ge H\left(X\mid Z\right)$$

当且仅当 $H(Y\mid X,Z)=0$ 成立即 Y 是 X 和 Z 的函数时等式成立。

(b)

根据 **Theorem 2.5.2** (*Chain rule for information*)

$$I(X,Y;Z)=I\left(X;Z\right)+I\left(Y;Z\mid X\right)\ge I\left(X;Z\right)$$

当且仅当 $I\left(Y;Z\mid X\right)=0$ 等式成立即 Y 和 Z 关于 X 是相互独立时等号成立。

(c)

$$
\begin{aligned}
&H(X,Y,Z)-H(X,Y)=H(X)+H(Y,Z\mid X)-H(X)-H(Y\mid X)\\
&=H(Y\mid X)+H(Z\mid X,Y)-H(Y\mid X)\\
&=H(Z\mid X,Y)=H(Z\mid X)-I(Z;X\mid Y)\\
&\le H(X,Z)-H(X)=H(X)+H(Z\mid X)-H(X)=H(Z\mid X)
\end{aligned}
$$

当且仅当 $I(Z;X\mid Y)=0$ 等式成立，即 Z 和 X 关于 Y 相互独立成立。

（d）根据 **Theorem 2.5.2** (*Chain rule for information*)(**未化简出来**)

$$I(X,Y;Z)=I(X;Z)+I(Y;Z\mid X)$$

$$I(X;Z\mid Y)+I(Z;Y)$$

2.30 最大熵。 设 X 是取非负整数值的随机变量,对固定的值 A>0,试求在约束条件

$$EX=\sum_{n=0}^{\infty} np(n)=A$$

下使得熵$H(X)$达到最大时的概率密度函数 p(x),并计算出 H(X)的最大值

解:

2.31 条件熵。在什么条件下有$H(X|\mathrm{g}(Y)) = H(X|\ Y)$?

解:

根据题意要证$H(X\mid G(Y))=H(X\mid Y)$ 直接证明左边等于右边似乎找不到联系，因此我们考虑尝试构造等价形式，观察左边右边等式条件熵有个共同随机变量 X。因此我们尝试构造：

$$H(X)-H(X\mid G(Y))=H(X)-H(X\mid Y)$$
$$\Leftrightarrow I(X;G(Y))=I(X;Y)$$

由此我们想到**Theorem 2.8.1** (*Data-processing inequality*) *If X → Y → Z, then*

$I(X;Y)\geq I(X;Z)$ *当且仅当等式成立时X → Y → Z,构成马尔科夫链，因此原问题当且仅当* $X\to G(Y)\to Y$ *构成马尔科夫链时上式成立。*

2.32 费诺。设(X, Y)的联合分布如右表:

X \ Y	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

设$\hat{X}(Y)$为 X 的估计量（基于 Y）,$P_e=\Pr\{\hat{X}(Y)\neq X\}$

（a） 试求最小误差概率估计量$\hat{X}(Y)$ 与相应的P_e

（b） 估计出啊该习题的费诺不等式，并与（a）中求得的值比较。

解:

2.33 费诺不等式。设$Pr(X=i)=\ p_i, i=1,2,\ldots,m$,且$p_1\geq p_2\geq p_3\geq\cdots\geq p_m$,那么 X 的最小误差概率估计量为$\hat{X}=1$,此时产生的误差概率为$P_e=1-P_1$。试在约束条件$1-P_1=P_e$。下最大化H(**p**),由此根据 H 求得 **p** 的取值范围。这也是无条件的费诺不等式。

解:

因为

$$H(P)=-\sum_{i=1}^{m} p_i \log p_i = -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i$$

要求$H(P_e)$ 的范围，必然需要构造$H(P_e)=-P_e\log P_e-(1-P_e)\log\left(1-P_e\right)$

因此重新化简：

$$\begin{aligned}
H(P)&=-\sum_{i=1}^{m} p_i \log p_i = -p_1 \log p_1 - P_e\sum_{i=2}^{m}\frac{p_i}{P_e}\log\frac{p_i}{P_e}-\log P_e\sum_{i=2}^{m}p_i \\
&=-\left(1-P_e\right)\log\left(1-P_e\right)-P_e\sum_{i=2}^{m}\frac{p_i}{P_e}\log\frac{p_i}{P_e}-\left(1-p_1\right)\log P_e \\
&=-\left(1-P_e\right)\log\left(1-P_e\right)-P_e\sum_{i=2}^{m}\frac{p_i}{P_e}\log\frac{p_i}{P_e}-P_e\log P_e \\
&=H\left(P_e\right)+P_eH\left(\frac{p_2}{P_e},\ldots,\frac{p_m}{P_e}\right)
\end{aligned}$$

根据熵得极值性，我们知道$H\left(\frac{p_2}{P_e},\ldots,\frac{p_m}{P_e}\right)\le\log(m-1)$当且仅当等概时等式成立，因此

$$\begin{aligned}
&H(P)\le H\left(P_e\right)+P_e\log(m-1) \\
&\Leftrightarrow \mathrm{H(X)}\le H\left(P_e\right)+P_e\log(m-1) \\
&\Rightarrow P_e\ge\frac{\mathrm{H(X)}-H\left(P_e\right)}{\log(m-1)}
\end{aligned}$$

2.34 初始条件熵。证明对任意的马尔可夫链，$H(X_0|X)$随 n 非减。

解：

证：根据题意，由随机变量的数列$X_1,\ldots,X_{n-1},X_n$ 构成的马尔科夫链，我们由此可想到：

Theorem 2.8.1 (*Data-processing inequality*) *If $X \to Y \to Z$, then*

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

对应到本题有：

$$\begin{aligned}
&I(X_0;X_{n-1})\ge I(X_0;X_n) \\
&\Leftrightarrow H(X_0)-H(X_0\mid X_{n-1})\ge H(X_0)-H(X_0\mid X_n) \\
&\Rightarrow H(X_0\mid X_n)\ge H(X_0\mid X_{n-1})
\end{aligned}$$

从而原命题得证。

2.35 相对熵是不对称的。设随机变量 X 有三个可能的结果$\{a,b,c\}$。考虑该随机变量上的两个分布(右表):

计算$H(p),H(q),D(p||q)$和$D(q||\ p)$,并验证在此情

字 符	$p(x)$	$q(x)$
a	$\frac{1}{2}$	$\frac{1}{3}$
b	$\frac{1}{4}$	$\frac{1}{3}$
c	$\frac{1}{4}$	$\frac{1}{3}$

况下$D(p||q) \neq D(q|| \ p)$。

解：
根据定义：

$$H(p) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4}$$
$$= 1.5 bits$$

$$H(q) = -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3}$$
$$= \log 3$$

$$D(p \| q) = \sum p(x)\log\frac{p(x)}{q(x)}$$
$$= \frac{1}{2}\log\frac{3}{2} + 2\frac{1}{4}\log\frac{3}{4} = \log 3 - \frac{3}{2}$$

$$D(q \| p) = \sum q(x)\log\frac{q(x)}{p(x)}$$
$$= \frac{1}{3}\log\frac{2}{3} + 2\frac{1}{3}\log\frac{4}{3} = \frac{5}{3} - \log 3$$

因此$D(p \| q) \neq D(q \| p)$

2.36 对称的相对熵。尽管如习题 2.35 所示，在一般情况下$D(p||q) \neq D(q||p)$,但也存在使等号成立的分布。请举出二元字母表上的两个分布 p 和 q,使得$D(p|| \ q) = D(q||p)$（除平凡情形 p=q 外）。

解：
二项分布应该是比较经典的一个符号要求的分布了。直接带入验证即可，很简单这里不再赘述。

2.37 相对熵。设三个随机变量 X、Y 和 Z 的联合概率密度函数为$p(x, y, z)$。联合分布和边际分布乘积之间的相对熵为

$$D\big(p(x,y,z)||p(x)p(y)p(z)\big) = E[\log\left(\frac{p(x,y,z)}{p(x)p(y)p(z)}\right)]$$

将上式用熵的形式展开。什么时候该相对熵为 0?

解：

$$D\left(p(x,y,z) \| p(x)p(y)p(z)\right) = E\left[\log\frac{p(x,y,z)}{p(x)p(y)p(z)}\right]$$
$$= \sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\left(\log\frac{p(x,y,z)}{p(x)p(y)p(z)}\right)$$

$$=\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z}p(x,y,z)\left(\log p(x,y,z)-\log p(x)-\log p(y)-\log p(z)\right)$$
$$=-H(X,Y,Z)+H(X)+H(Y)+H(Z)$$

要使得$D\left(p(x,y,z)\,\|\,p(x)p(y)p(z)\right)=0$

当且仅当$\frac{p(x,y,z)}{p(x)p(y)p(z)}\Rightarrow p(x,y,z)=p(x)p(y)p(z)$即 X,Y,Z 相互独立。

2.38 问题的值。设$X\sim p(x)$,$x=1,2,\dots,m$,给定一个集合$S\subseteq\{1,2,\dots,m\}$。是否当 X∈S 时，得到的答案为

$$Y=\begin{cases}1 & ,X\in S\\ 0 & ,X\notin S\end{cases}$$

假定$\Pr\{X\in S\}=\alpha$,试求不确定度的缩减量$H(X)-H(X|Y)$。显然，给定α,任何集合 S 的表现与其他的集合是一样的。

解：

由$P_r\{X\in S\}=\alpha,Y=\begin{cases}1 & X\in S\\ 0 & X\notin S\end{cases}$我们有

$$\Rightarrow\begin{cases}p(y=1)=\alpha\\ p(y=0)=1-\alpha\end{cases}$$

$$H(X)-H(X|Y)=I(X;Y)$$
又$I(X;Y)=H(Y)-H(Y|X)$
$$=H(\alpha)-H(Y|X)$$

又因为 Y 本质上是关于 X 的函数，因此$H(Y|X)=0$

所以：
$$\begin{aligned}H(X)-H(X|Y)&=I(X;Y)=H(Y)-H(Y|X)\\&=H(\alpha)-H(Y|X)\\&=H(\alpha)\end{aligned}$$

2.39 熵与两两独立。设 X，Y 和 Z 为三个服从 Bemouli($\frac{1}{2}$)的二元随机变量，且两两相互独立，即$I(X;Y)=I(X;Z)=\ I(Y;Z)=0$。

(a)在上述约束条件下。$H(X,Y,Z)$的最小值是多少?

(b)举出达到这个最小值时的例子。

解：

(a)

$$
\begin{aligned}
H(X,Y,Z) &= H(X)+H(Y,Z\mid X) \\
&= H(X)+H(Y\mid X)+H(Z\mid X,Y)
\end{aligned}
$$

又因为·X,Y,Z 两两相互独立，因此 $H(Y\mid X)=H(Y)$

因此上式可放缩为：

$$
\begin{aligned}
H(X,Y,Z) &= H(X)+H(Y,Z\mid X) \\
&= H(X)+H(Y\mid X)+H(Z\mid X,Y) \ge H(X)+H(Y) \\
&= -2(2\frac{1}{2}\log\frac{1}{2}) = 2
\end{aligned}
$$

因此最小值为 2.。

（b）

2.40 离散熵。设 X 和 Y 为两个独立且取整数值的随机变量。设 X 在{1,2,...,8}上均匀分布，$\Pr\{Y=k\}=2^{-k}, k=1,2,3\ldots$。

（a）求$H(X)$。

（b）求$H(Y)$。

（c）求$H(X+Y,X-Y)$。

解：

(a)

方法一： 根据熵得定义：

$H(X)=-\sum_{k=1}^{8} p_k \log p_k$ 因为等概，因此有

$$H(X)=-8\frac{1}{8}\log\frac{1}{8}=3$$

方法二：根据熵的**极值定理：**

$H(P_1,P_2,\ldots,P_q)\le \log q$ **当** $P_1=P_2\ldots=P_q=1/q$ 时，信源具有最大熵，即等式成立。

因此 $H(X)=\log 8=3$

（b） 根据熵得定义：

$$H(Y)=-\sum_{k=1}^{8} p_k \log p_k = -\sum_{k=1}^{8} 2^{-k}\log 2^{-k} = \sum_{k=1}^{8} k2^{-k} = 2$$

(c) $H(X+Y,X-Y)=H(X+Y)+H\left(X-Y\mid X+Y\right)$

令 $Z=X+Y, U=X-Y$

$$H(X+Y,X-Y)=H(X+Y)+H\left(X-Y\mid X+Y\right)$$
$$=H(Z,U)=H(Z)+H\left(U\mid Z\right)$$

因此我们只需要（分类讨论情况是不是有点多？？有问题。）

2.41 随机问题，要判别随机目标 X~p(x)，问题 Q~r(q)随机地提问，结果产生的确定答案$A=A(x,q)\in\{a_1,a_2,\ldots\}$。假定 X,Q 相互独立。于是$I(X;Q,A)$ 为由问题-答案对（Q，A）之后 X 剩下的不确定性。

（a）证明$I(X;Q,A)=H(A|Q)$，并予以解释。

（b）现在假定有两个 i.i.d 的问题$Q_1,Q_2\sim r(q)$提出，其答案分别为A_1，A_2 证明$I(X;Q,A_1,Q_2,A_2)\le 2I(X;Q,A_1)$在此意义下，说明两个问题不比单个问题两次的效果更差。

解：

（a）

$$I(X;Q,A)=H(X)-H(X\mid Q,A)$$
$$=H(Q,A)-H(Q,A\mid X)$$
$$=H(Q)+H(A\mid Q)-H(Q\mid X)-H(A\mid Q,X)$$

根据题意我们知道 X 和 Q 相互独立，因此$H(Q\mid X)=H(Q)$，又因为 Q 随机提问后，A 随之确定，即 A 由 X 和 Q 决定，因此$H(A\mid Q,X)=0$。综上们有：

$$I(X;Q,A)=H(X)-H(X\mid Q,A)$$
$$=H(Q,A)-H(Q,A\mid X)$$
$$=H(Q)+H(A\mid Q)-H(Q\mid X)-H(A\mid Q,X)$$
$$=H(Q)+H(A\mid Q)-H(Q)-0$$
$$=H(A\mid Q)$$

（b）

根据 **Theorem 2.5.2** (*Chain rule for information*)，我们有：

$$I(X;Q_1,A_1,Q_2,A_2)=I(X;Q_1)+I(X;A_1\mid Q_1)+I\left(X;Q_2\mid A_1,Q_1\right)+I\left(X;Q_2\mid A_1,Q_1,Q_2\right)$$

又因为 X 和Q_1相互独立，　$I(X;Q_1)=0$，因此

$$\begin{aligned}I(X;Q_1,A_1,Q_2,A_2)&=I(X;Q_1)+I(X;A_1\mid Q_1)+I\left(X;Q_2\mid A_1,Q_1\right)+I\left(X;A_2\mid A_1,Q_1,Q_2\right)\\&=I(X;A_1\mid Q_1)+I\left(X;Q_2\mid A_1,Q_1\right)+I\left(X;Q_2\mid A_1,Q_1,Q_2\right)\\&=H(A_1\mid Q_1)-H(A_1\mid X,Q_1)+H(Q_2\mid A_1,Q_1)-H(Q_2\mid X,A_1,Q_1)+H(A_2\mid A_1,Q_1,Q_2)-H(A_2\mid X,A_1,Q_1,Q_2)\end{aligned}$$

而由已知条件知道，Q_1与A_1,Q_1相互独立，且A_1由X,Q_1，所确定，因此Q_2与A_1,Q_1,Q_2相互独立，且A_2由X,Q_2，所确定，因此

$$H(A_1|X,Q_1)=0$$

$$H(Q_2|A_1,Q_1)=H(Q_2|X,A_1,Q_1)=H(Q_2)$$

$$H(A_2|X,A_1,Q_1,Q_2)=0$$

从而

$$\begin{aligned}I(X;Q_1,A_1,Q_2,A_2)&=I(X;Q_1)+I(X;A_1|Q_1)+I\left(X;Q_2|A_1,Q_1\right)+I\left(X;Q_2|A_1,Q_1,Q_2\right)\\&=H(A_1|Q_1)-H(A_1|X,Q_1)+I\left(X;Q_2|A_1,Q_1\right)+I\left(X;Q_2|A_1,Q_1,Q_2\right)\\&=H(A_1|Q_1)-H(A_1|X,Q_1)+H(Q_2|A_1,Q_1)-H(Q_2|X,A_1,Q_1)+H(A_2|A_1,Q_1,Q_2)-H(A_2|X,A_1,Q_1,Q_2)\\&=H(A_1|Q_1)+H(Q_2)-H(Q_2)+H(A_2|A_1,Q_1,Q_2)\\&=H(A_1|Q_1)+H(A_2|A_1,Q_1,Q_2)\end{aligned}$$

又因为条件熵使得熵减小，因此$H(A_2|A_1,Q_1,Q_2)\le H(A_2|Q_2)$，再根据（a）中结论有

$$H(A_1|Q_1)=I(X;Q_1,A_1)$$

$$H(A_2|Q_2)=I(X;Q_2,A_2)$$

$$\begin{aligned}I(X;Q_1,A_1,Q_2,A_2)&=I(X;Q_1)+I(X;A_1|Q_1)+I\left(X;Q_2|A_1,Q_1\right)+I\left(X;Q_2|A_1,Q_1,Q_2\right)\\&=I(X;A_1|Q_1)+I\left(X;Q_2|A_1,Q_1\right)+I\left(X;Q_2|A_1,Q_1,Q_2\right)\\&=H(A_1|Q_1)-H(A_1|X,Q_1)+H(Q_2|A_1,Q_1)-H(Q_2|X,A_1,Q_1)+H(A_2|A_1,Q_1,Q_2)-H(A_2|X,A_1,Q_1,Q_2)\\&=H(A_1|Q_1)-H(A_1|X,Q_1)+H(Q_2)-H(Q_2)+H(A_2|A_1,Q_1,Q_2)\\&=H(A_1|Q_1)+H(A_2|A_1,Q_1,Q_2)\\&\le H(A_1|Q_1)+H(A_2|Q_2)\\&=I(X;Q_1,A_1)+I(X;Q_2,A_2)\end{aligned}$$

又因为Q_1和Q_2同分布，因此

$$I(X;Q_1,A_1)=I(X;Q_2,A_2)$$

$$\begin{aligned}I(X;Q_1,A_1,Q_2,A_2)&=I(X;Q_1)+I(X;A_1|Q_1)+I\left(X;Q_2|A_1,Q_1\right)+I\left(X;Q_2|A_1,Q_1,Q_2\right)\\&\le 2I(X;Q_1,A_1)\end{aligned}$$

综上原命题得证。

2.42 不等式。下列不等式在一般情况下是“≥"、“="还是“≤"关系?请将每个不等式用“≥"、“="或“≤"标出各自的正确关系。

(a) $H(5X)$与$H(X)$。

(b) $I(g(X);Y)$与$I(X;Y)$。

(c) $H(X_0|X_{-1})$与$H(X_0|X_{-1},X_1)$。

(d) $H(X,Y)/(H(X)+H(Y))$与 1。

解：

(a)

(b)

令 $Z=g(X)$ 则 $X\to Y\to g(X)$ 构成马尔科夫链，因此，根据 **Theorem 2.8.1** (*Data-processing inequality*) *If X → Y → Z, then* $I(X;Y)\geq I(X;Z)$*,从而有:*

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

(c)

根据 **Theorem 2.6.5** (*Conditioning reduces entropy*)(*Information can't hurt*)

$H(X|Y)\leq H(X)$ ，从而有 $H(X_0|X_{-1})\geq H(X_0|X_{-1},X_1)$

(d)

因为 $H(X,Y)=H(X)+H(Y|X)\leq^a H(X)+H(Y)$ 其中a利用了**Theorem 2.6.5**

因此

$$H(X,Y)\leq H(X)+H(Y)\Rightarrow\frac{H(X,Y)}{H(X)+H(Y)}\leq 1$$

2.43 正面和反面的互信息。

(a)考虑抛掷一枚均匀硬币。硬币出现正面和反面的互信息是多少?

(b)如果我们掷一颗有 6 面的均匀骰子，那么顶面和前面(经常面对你的那个侧面)出现的互信息又是多少?

解:

(a)

假设正面记为A，反面记为B，则有

$$I(A;B)=H(A)-H(A|B)$$

(b)

2.44 纯随机性。假定用一枚具有三面的硬币来产生均匀硬币抛掷过程。设硬币 X 的概率密度函数为

$$X=\begin{cases}A, & p_A\\ B, & p_B\\ C, & p_C\end{cases}$$

其中P_A、p_B和p_C未知。

(a)如何通过两个独立的抛掷X_1和X_2产生(如果可行)一个 Bernoulli($\frac{1}{2}$)随机变量 Z?

(b)生成的最大均匀二进制序列的数量的期望数是多少?

解：

2.45有限熵。证明:对于离散随机变量$X \in \{1,2,...\}$,如果$ElogX < \infty$,则$H(X) < \infty$。
证明:

2.46 熵的公理化定义(较难)。如果为度量信息而假定某些公理,将不得不使用如熵那样的对数度量。香农利用这点确保了熵的最初定义的合理性。在本书中，我们更多依赖于熵的其他性质而非公理化推导来确保它的使用价值。下面这个题比起本节的其他习题要困难多。

若对称函数序列$H(p_1,p_2,...,p_m)$满足下列性质:

●标准化:$H_2(\frac{1}{2},\frac{1}{2})$=1,

●连续性:$H_2(p,1-p)$为 p 的连续函数,

●组合法则:$H_m(p_1,p_2,...p_m) = H_{m-1}(p_1+p_2,p_3,...,p_m) + (p1+p_2)\cdot H_2(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2})$

证明H_m必定具有如下形式:

$$H_m(p_1,p_2,...,p_m) = -\sum_{i=1}^{m} p_i \log p_i\,, \qquad m = 2,3,...$$

还有许多不同的公理化表示方式可以导出熵的相同定义。例如，可参见 Csiszor 和 Korner[149]。

解：

2.47 分类错误文件的熵。一副扑克牌共有 n 张，顺序依次为 1,2,...,n。现在从这副扑克中随机地抽出一张牌，然后再随机地将其放回。这样，熵为多少?

解：

2.48 序列长度。序列的长度含有序列内容的多少信息?假定考虑 Bernoulli($\frac{1}{2}$)过程$\{X_i\}$,当第一个 1 出现时，过程停止。设 N 表示这个停时。因此，X^N为所有有限长的二元序列集合$\{0,1\}^* = \{0,1,00,01,10,11,000,...\}$中的一个元素。

(a) 求$I(N;X^N)$

(b) 求$H(X^N|N)$

(c) 求$H(X^N)$

现考虑一个不同的停时。仍假定X_i~Bernoulli($\frac{1}{2}$)，但过程在时刻$N = 6$停止的概率为 $\frac{1}{3}$

在时刻$N = 12$停止的概率为$\frac{2}{3}$设该停时独立于序列$X_1,X_2,...,X_{12}$。

(d) 求$I(N;X^N)$

（e）求$H(X^N|N)$

（f）求$H(X^N)$

解：

（a） $I(X^N;N)=H(N)-H(N|X^N)$

又因为X^N是关于N的函数，因此$H(N|X^N)=0$，因此

$I(X^N;N)=H(N)-H(N|X^N)=H(N)-0=H(N)$

又因为$P(N=n)=pq^{n-1}$ 几何概率分布，根据本章习题2.1，$H(N)=\sum_{n=1}^{\infty}n2^{-n}=2bits$

(b)

因为给定N，X^N 随之确定，X^N是关于N的函数，因此$H(X^N|N)=0$

(c)

要求$H(X^N)$ 首先想到（b）中出现的$H(X^N|N)$，再根据（a）中已求得的

$$
\begin{aligned}
&I(X^N;N)=H(X^N)-H(X^N|N)\\
&\Rightarrow H(X^N)=I(X^N;N)+H(X^N|N)\\
&\Rightarrow H(X^N)=I(X^N;N)=2bits
\end{aligned}
$$

(d)

根据（a）$I(X^N;N)=H(N)-H(N|X^N)=H(N)-0=H(N)$

又因为N=6停止概率为$\frac{1}{3}$ ，N=12，停止概率为$\frac{2}{3}$ ，因此

$I(X^N;N)=\frac{1}{3}H(N)+\frac{2}{3}H(N)=1bits$

(e)

这里会有两次停止的概率，即N=6停止概率为$\frac{1}{3}$ ，N=12，停止概率为$\frac{2}{3}$，N的取值不再是一一对应了，而是概率出现了。因此

$H(X^N|N)=\frac{1}{3}H(X^6|N=6)+\frac{2}{3}H(X^{12}|N=12)$

(f)

由**（c）**中知

$$
\begin{aligned}
&I(X^N;N)=H(X^N)-H(X^N|N)\\
&\Rightarrow H(X^N)=I(X^N;N)+H(X^N|N)
\end{aligned}
$$