EE/Ma 126a Information Theory - Homework Set #2

Ling Li, ling@cs.caltech.edu

October 10, 2000

2.1 Venn diagrams

By definitions or properties of mutual information and the chain rule for entropy, we know

$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) & (1) \\ I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) & (2) \\ I(Z;X) &= H(X) - H(X|Z) & (3) \\ I(Y;Z) &= H(Z) - H(Z|Y) & (4) \\ H(X,Y,Z) &= H(Y) + H(Z|Y) + H(X|Y,Z) & (5) \end{aligned}$$

(1)-(2)-(3)-(4)-(5) is what we want:

$$\boxed{I(X;Y;Z) = H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X)} \quad (6)$$

We also have $I(X;Y) = H(X) + H(Y) - H(X,Y)$, and the other two similar equations for $I(Y;Z)$ and $I(Z;X)$. Plug them into (6), that is

$$\boxed{I(X;Y;Z) = H(X,Y,Z) + H(X) + H(Y) + H(Z) - H(X,Y) - H(Y,Z) - H(Z,X)} \quad (7)$$

Let X, Y and Z have the following joint distribution:

Z \ X,Y	$0,0$	$0,1$	$1,0$	$1,1$
0	1/4	0	0	1/4
1	0	1/4	1/4	0

Thus $H(X) = H(Y) = H(Z) = 1$, $H(X,Y) = H(Y,Z) = H(Z,X) = 2$, and $H(X,Y,Z) = 2$. From (7) we have

$$\boxed{I(X;Y;Z) = -1}$$

2.2 Drawing with and without replacement

$\boxed{\text{The entropy with replacement will be higher.}}$

Let random variable X_i denotes the color of the ith ball drawn from the urn. The entropy of drawing $k \geq 2$ balls is

$$\begin{aligned} H(X_1, X_2, \ldots, X_k) &= H(X_1) + H(X_2|X_1) + \cdots + H(X_k|X_{k-1}, \ldots, X_1) && (8) \\ &\leq H(X_1) + H(X_2) + \cdots + H(X_k) && (9) \end{aligned}$$

with equality iff $X_1, X_2, \ldots, X_k$ are independent. If the drawing is with replacement, the entropy will be (9). Otherwise, $X_1, X_2, \ldots, X_k$ are not independent and the entropy is something like (8), always less than the previous one.

2.3 Entropy of a sum

(a) When X is specified, $Z = X + Y$ is a one-to-one function of Y, so $\boxed{H(Z|X) = H(Y|X)}$. If X, Y are independent, we have

$$\boxed{H(Y) = H(Y|X) = H(Z|X) \leq H(Z)}$$

Since Z is symmetric in X and Y, we also have $\boxed{H(X) \leq H(Z)}$.

(b) Let (X, Y) have the following joint distribution:

Y \ X	0	1
0	0	1/2
1	1/2	0

The distribution of Z is $p(1) = 1$. Thus $\boxed{H(X) = H(Y) = 1}$ and $\boxed{H(Z) = 0}$. In this case, $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) (X, Y) can be considered to be a single vector-valued random variable, and Z is the function of (X, Y). Thus

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

with the first equality iff $\boxed{Z = X + Y \text{ is a one-to-one mapping from } (X,Y) \text{ to } Z}$, and with the second equality iff $\boxed{X,\ Y \text{ are independent}}$. Thus $H(Z) = H(X) + H(Y)$ iff X, Y are indepent and $Z = X + Y$ is a one-to-one mapping from (X, Y) to Z.

2.4 Entropy of a disjoint mixture

(a) $\mathcal{X}_1 \bigcap \mathcal{X}_2 = \emptyset$,

$$p(x) = \begin{cases} \alpha p_1(x), & x \in \mathcal{X}_1; \\ (1-\alpha) p_2(x), & x \in \mathcal{X}_2. \end{cases}$$

Thus

$$\begin{aligned}
H(X) &= -\sum_{x\in\mathcal{X}_1\bigcup\mathcal{X}_2} p(x)\log p(x) \\
&= -\sum_{x\in\mathcal{X}_1}\alpha p_1(x)\log\left[\alpha p_1(x)\right] - \sum_{x\in\mathcal{X}_2}(1-\alpha)p_2(x)\log\left[(1-\alpha)p_2(x)\right] \\
&= -\alpha\sum_{x\in\mathcal{X}_1} p_1(x)\left[\log\alpha + \log p_1(x)\right] - (1-\alpha)\sum_{x\in\mathcal{X}_2} p_2(x)\left[\log(1-\alpha)+\log p_2(x)\right] \\
&= -\alpha\sum_{x\in\mathcal{X}_1} p_1(x)\log p_1(x) - (1-\alpha)\sum_{x\in\mathcal{X}_2} p_2(x)\log p_2(x) - \alpha\log\alpha - (1-\alpha)\log(1-\alpha) \\
&= \boxed{\alpha H(X_1) + (1-\alpha)H(X_2) - \alpha\log\alpha - (1-\alpha)\log(1-\alpha)}
\end{aligned}$$

(b) Let $\dfrac{\partial H(X)}{\partial\alpha} = H(X_1) - H(X_2) - \log\dfrac{\alpha}{1-\alpha} = 0$, we get

$$\boxed{\alpha = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}}$$

With this α, $H(X)$ reaches the maximum

$$\max\{H(X)\} = \log\left(2^{H(X_1)} + 2^{H(X_2)}\right)$$

since $\dfrac{\partial^2 H(X)}{\partial\alpha^2} = -\dfrac{1}{\alpha(1-\alpha)}$ is alwalys less than 0. That is to say,

$$\boxed{2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}} \tag{10}$$

If we combine the effective alphabets for $\mathcal{X}_1$ and $\mathcal{X}_2$, the combined one will be an alphabets for $\mathcal{X} = \mathcal{X}_1\bigcup\mathcal{X}_2$, but it will not be more 'effective' than the effective alphabets for $\mathcal{X}$. That is, the sum of the effective alphabet sizes of $\mathcal{X}_1$ and $\mathcal{X}_2$ will not be less than the size of the effecitive alphabet of $\mathcal{X}$, which is just (10).

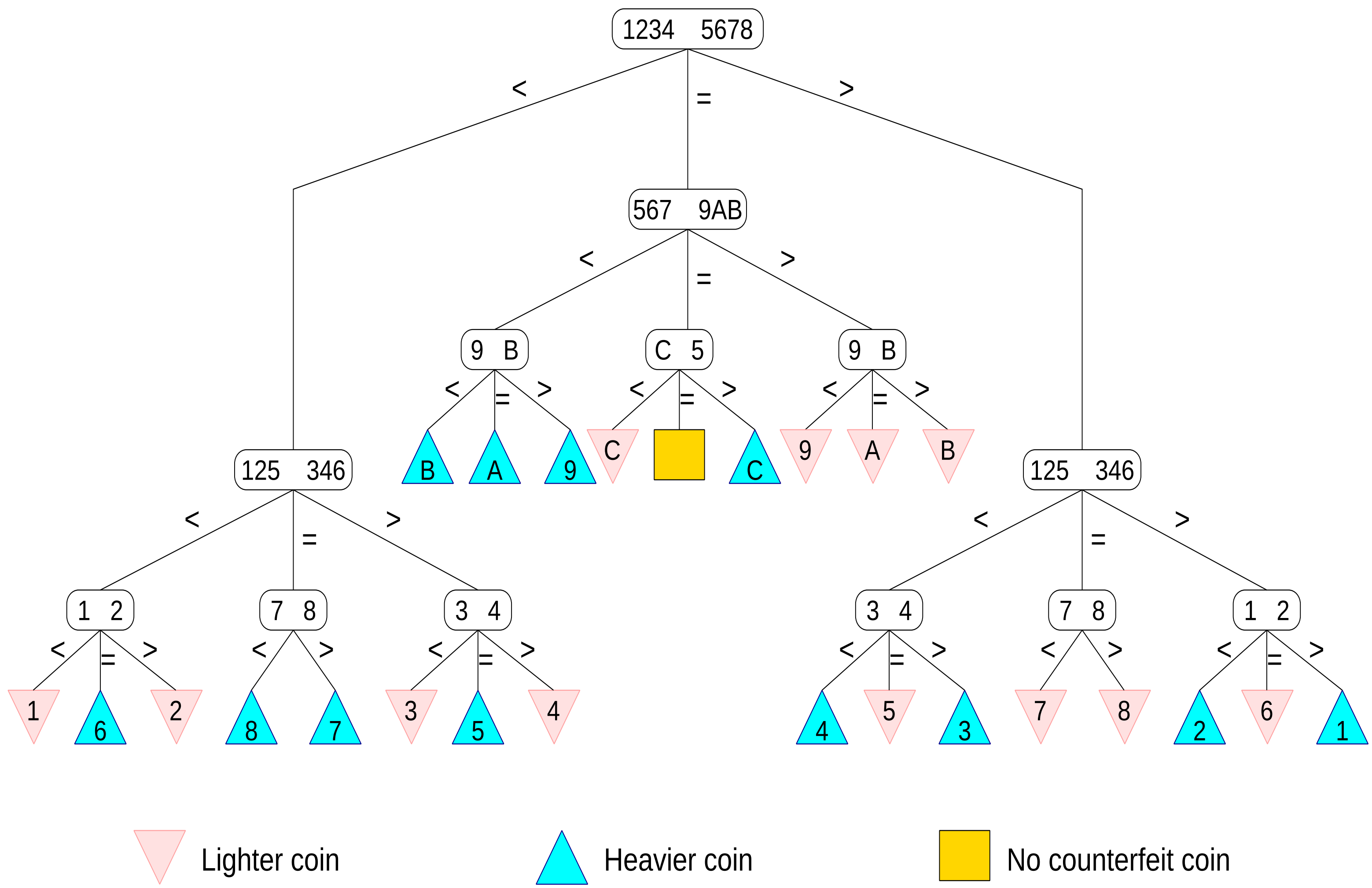
2.5 Coin weighing

(a) There're at most 3 possible results that 1 weighing can give: left lighter than right, balance, or left heavier than right. There're at most 3×3 possible results that 2 weighings can give. For k weighings, there're at most 3^k possible results. If we want to find the counterfeit coin and declare it to be heavier or lighter, or declare that there's no counterfet coin, in no more than k weighings, the number of different situations should be no more than 3^k. For n coins, the number of different possible situations is $(2n+1)$ since each coin can be heavier or lighter, or no one is counterfeit. In order to distinguish all these situations, we have

$$3^k \geq 2n+1$$

Thus an upper bound of n is

$$\boxed{n \leq \frac{3^k - 1}{2}}$$



(b) Number $1, 2, \ldots, A, B, C$ denote different coins.