Exercise 10.3.2

Here we extend our discussion in Exercise 10.3.1 to $n-$dimension entropy. We still have the entropy of a discrete random variable X as

$$H(X) = -\sum_{x} p_X(x)\log\left[p_X(x)\right] \tag{1}$$

The conditional entropy of multiple discrete random variable X_i is defined by

$$H(X_k|X_1,\ldots,X_{k-1}) = -\sum_{x_1,\ldots,x_k} p_{X_1,\ldots,X_k}(x_1,\ldots,x_k)\log\left[p_{X_k|X_1,\ldots,X_{k-1}}(x_k|x_1,\ldots,x_{k-1})\right] \tag{2}$$

The joint entropy of multiple discrete random variable X_i is defined by

$$H(X_1,\ldots,X_k) = -\sum_{x_1,\ldots,x_k} p_{X_1,\ldots,X_k}(x_1,\ldots,x_k)\log\left[p_{X_1,\ldots,X_k}(x_1,\ldots,x_k)\right] \tag{3}$$

We still have the multiplicative probability relation for any k,

$$p_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = p_{X_k|X_1,\ldots,X_{k-1}}(x_k|x_1,\ldots,x_{k-1})p_{X_1,\ldots,X_{k-1}}(x_1,\ldots,x_{k-1}) \tag{4}$$

From eq. (4) we have for eq. (3),

$$\begin{aligned} H(X_1,\ldots,X_n) &= -\sum_{x_1,\ldots,x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\log\left[p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\right] \\ &= -\sum_{x_1,\ldots,x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\log\left[p_{X_n|X_1,\ldots,X_{n-1}}(x_n|x_1,\ldots,x_{n-1})\right] \\ &\quad -\sum_{x_1,\ldots,x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\log\left[p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})\right] \end{aligned} \tag{5}$$

If we sum up x_n in the second term of the last identity in eq. (5), and compare eq. (5) with eq. $(1)-(3)$, we will have

$$H(X_1,\ldots,X_n) = H(X_n|X_1,\ldots,X_{n-1}) + H(X_1,\ldots,X_{n-1}) \tag{6}$$

For $H(X_1,\ldots,X_{n-1})$, we can use similar method in eq. $(4)-(5)$ and get the joint probability as

$$p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) = p_{X_{n-1}|X_1,\ldots,X_{n-2}}(x_{n-1}|x_1,\ldots,x_{n-2})p_{X_1,\ldots,X_{n-2}}(x_1,\ldots,x_{n-2}) \tag{7}$$

and also

$$\begin{aligned} H(X_1,\ldots,X_{n-1}) &= -\sum_{x_1,\ldots,x_{n-1}} p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})\log\left[p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})\right] \\ &= -\sum_{x_1,\ldots,x_{n-1}} p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})\log\left[p_{X_{n-1}|X_1,\ldots,X_{n-2}}(x_{n-1}|x_1,\ldots,x_{n-2})\right] \\ &\quad -\sum_{x_1,\ldots,x_{n-1}} p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})\log\left[p_{X_1,\ldots,X_{n-2}}(x_1,\ldots,x_{n-2})\right] \\ &= H(X_{n-1}|X_1,\ldots,X_{n-2}) + H(X_1,\ldots,X_{n-2}) \end{aligned} \tag{8}$$

So we have for $H(X_1,\ldots,X_n)$ from eq. (8) as

$$H(X_1,\dots,X_n) = H(X_n|X_1,\dots,X_{n-1}) + H(X_{n-1}|X_1,\dots,X_{n-2}) + H(X_1,\dots,X_{n-2}) \tag{9}$$

Repeat the same method from eq. (7) − (9), we will eliminate $H(X_1,\dots,X_k)$ terms and finally have

$$\begin{aligned} H(X_1,\dots,X_n) =& H(X_n|X_1,\dots,X_{n-1}) + H(X_{n-1}|X_1,\dots,X_{n-2}) \\ &+ H(X_{n-2}|X_1,\dots,X_{n-3}) + \cdots + H(X_2|X_1) + H(X_1) \end{aligned} \tag{10}$$