Exercise 10.3.1

The entropy of a discrete random variable X is defined by

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

The conditional entropy of discrete random variable X and Y is defined by

$$H(X|Y) = -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X|Y}(x|y) \right] \tag{2}$$

The joint entropy of discrete random variable X and Y is defined by

$$H(X,Y) = -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X,Y}(x,y) \right] \tag{3}$$

From the definition of conditional probability, we have $p_{X,Y}(x,y) = p_{Y|X}(y|x) p_X(x)$, then eq. (3) becomes

$$\begin{aligned} H(X,Y) &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X,Y}(x,y) \right] \\ &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{Y|X}(y|x) p_X(x) \right] \\ &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{Y|X}(y|x) \right] - \sum_{x,y} p_{X,Y}(x,y) \log \left[p_X(x) \right] \end{aligned} \tag{4}$$

Compare with eq. $(1) - (3)$, the first term in the last identity of eq. (4) is $H(Y|X)$, and the second term in the last identity of eq. (4) is $H(X)$ if we sum up over y. Thus, we have

$$H(X,Y) = H(X) + H(Y|X) \tag{5}$$

Similarly, we can re-write eq. (4) with $p_{X,Y}(x,y) = p_{X|Y}(x,y) p_Y(y)$ and obtain

$$\begin{aligned} H(X,Y) &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X,Y}(x,y) \right] \\ &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X|Y}(x|y) p_Y(y) \right] \\ &= -\sum_{x,y} p_{X,Y}(x,y) \log \left[p_{X|Y}(x|y) \right] - \sum_{x,y} p_{X,Y}(x,y) \log \left[p_Y(y) \right] \end{aligned} \tag{6}$$

Compare with eq. $(1) - (3)$, the first term in the last identity of eq. (6) is $H(Y|X)$, and the second term in the last identity of eq. (6) is $H(Y)$ if we sum up over x. Thus, we have

$$H(X,Y) = H(Y) + H(X|Y) \tag{7}$$