Course «Information theory». Solution of an exercise from the homework.

Exercise 5.1. Let (α, β, γ) be a triple of jointly distributed random variables. For every value c_k of γ we have a conditional distribution of probabilities of the values (α, β) (conditional on $\gamma = c_k$), $p_{ij} = \text{Prob}[\alpha = a_i \ \& \ \beta = b_j \,|\, \gamma = c_k]$. For this conditional distribution we introduce the quantity that is called the mutual information between α and β; we denote this quantity $I(\alpha : \beta \,|\, \gamma = c_k)$.

Let us define the *conditional mutual information* between α and β given γ as

$$I(\alpha : \beta \,|\, \gamma) := \sum_k \text{Prob}[\gamma = c_k] \cdot I(\alpha : \beta \,|\, \gamma = c_k)$$

Prove that

$$\begin{array}{rcl} I(\alpha : \beta \,|\, \gamma) & \stackrel{(1)}{=} & H(\beta \,|\, \gamma) - H(\beta \,|\, \alpha, \gamma) \\ & \stackrel{(2)}{=} & H(\alpha \,|\, \gamma) - H(\alpha \,|\, \beta, \gamma) \\ & \stackrel{(3)}{=} & H(\alpha \,|\, \gamma) + H(\beta \,|\, \gamma) - H(\alpha, \beta \,|\, \gamma) \\ & \stackrel{(4)}{=} & H(\alpha, \gamma) + H(\beta, \gamma) - H(\alpha, \beta, \gamma) - H(\gamma). \end{array}$$

Solution.

Eq. (3) We know that for *every* distribution (X, Y)

$$I(X : Y) = H(X) + H(Y) - H(X, Y) = H(X) - H(X \,|\, Y) = H(Y) - H(Y \,|\, X).$$

In particular, we can apply these equations for the joint distribution of α and β, *conditional on each value of* γ.

$$\begin{array}{rcl} I(\alpha : \beta \,|\, \gamma) & := & \sum\limits_k \text{Prob}[\gamma = c_k] \cdot I(\alpha : \beta \,|\, \gamma = c_k) \\ & = & \sum\limits_k \text{Prob}[\gamma = c_k] \cdot \big(H(\alpha \,|\, \gamma = c_k) + H(\beta \,|\, \gamma = c_k) - H(\alpha, \beta \,|\, \gamma = c_k)\big) \\ & = & \sum\limits_k \text{Prob}[\gamma = c_k] \cdot H(\alpha \,|\, \gamma = c_k) + \sum\limits_k \text{Prob}[\gamma = c_k] \cdot H(\beta \,|\, \gamma = c_k) \\ & & - \sum\limits_k \text{Prob}[\gamma = c_k] \cdot H(\alpha, \beta \,|\, \gamma = c_k) \\ & = & H(\alpha \,|\, \gamma) + H(\beta \,|\, \gamma) - H(\alpha, \beta \,|\, \gamma). \end{array}$$

Eq. (4) We know that $H(\alpha \,|\, \gamma) = H(\alpha, \gamma) - H(\gamma)$, $H(\beta \,|\, \gamma) = H(\beta, \gamma) - H(\gamma)$, and $H(\alpha, \beta \,|\, \gamma) = H(\alpha, \beta, \gamma) - H(\gamma)$. Combining these equations with Eq. (3) we obtain

$$\begin{array}{rcl} I(\alpha : \beta \,|\, \gamma) & = & H(\alpha \,|\, \gamma) + H(\beta \,|\, \gamma) - H(\alpha, \beta \,|\, \gamma) \\ & = & H(\alpha, \gamma) - H(\gamma) + H(\beta, \gamma) - H(\gamma) - H(\alpha, \beta, \gamma) + H(\gamma) \\ & = & H(\alpha, \gamma) + H(\beta, \gamma) - H(\alpha, \beta, \gamma) - H(\gamma) \end{array}$$

Eq. (1) is very similar to Eq. (3) : at first we notice that

$$\begin{array}{rcl} I(\alpha : \beta \,|\, \gamma) & := & \sum\limits_k \text{Prob}[\gamma = c_k] \cdot I(\alpha : \beta \,|\, \gamma = c_k) \\ & = & \sum\limits_k \text{Prob}[\gamma = c_k] \cdot \big(H(\beta \,|\, \gamma = c_k) + H(\beta \,|\, \alpha, \gamma = c_k)\big) \quad (*) \end{array}$$

By definition of conditional entropy,

$$H(Y \,|\, X) := \sum_i \text{Prob}[X = a_i] \cdot H(Y \,|\, X = a_i).$$

In particular, for the distribution of (α, β) conditional on $\gamma = c_k$ we have the following formula of the conditional entropy :

$$H(\beta \,|\, \alpha, \gamma = c_k) := \sum_i \text{Prob}[\alpha = a_i \,|\, \gamma = c_k] \cdot H(\beta \,|\, \alpha = a_i, \gamma = c_k).$$

Therefore, (*) can be rewritten as

$$\begin{aligned} & \sum_k \text{Prob}[\gamma = c_k] \cdot H(\beta \,|\, \gamma = c_k) + \sum_{i,k} \text{Prob}[\gamma = c_k] \cdot \text{Prob}[\alpha = a_i \,|\, \gamma = c_k] \cdot H(\beta \,|\, \alpha = a_i, \gamma = c_k) \\ = & \sum_k \text{Prob}[\gamma = c_k] \cdot H(\beta \,|\, \gamma = c_k) + \sum_{i,k} \text{Prob}[\alpha = a_i, \gamma = c_k] \cdot H(\beta \,|\, \alpha = a_i, \gamma = c_k) \end{aligned}$$

The last line can be rewritten (again, just by definition of the conditional entropy) as $H(\beta \,|\, \gamma) + H(\beta \,|\, \alpha, \gamma)$.

Eq. (2) is analogous to Eq. (1).