Information Theory

Exercise Sheet 3

Exercise 1 (About the weak law of large number). *Our aim in this exercise is to prove the weak law of large number. Recall Markov Inequality, for any positive random variable* $\mathbf{X}$*,*

$$\forall t \geq 0,\ \mathbb{P}(\mathbf{X} \geq t) \leq \frac{\mathbb{E}(\mathbf{X})}{t}$$

1. *Prove the Bienaymé-Tchebychev inequality, for any* $\alpha \geq 0$*,*

$$\mathbb{P}\left(|\mathbf{X} - \mathbb{E}(\mathbf{X})| \geq \alpha\right) \leq \frac{\mathbb{V}(\mathbf{X})}{\alpha^2}$$

2. *Prove the weak law of large number.*

3. *Prove the weak law of large number with dependency as seen during the lecture.*

Exercise 2 (Characterization of Markov chains with uniform stationary distribution). *Show that the uniform distribution is a stationary distribution of a Markov Chain of order one if and only if its probability transition matrix* $\mathbf{P} = (p(x,y))$ *is doubly stochastic, i.e.,*

$$\sum_y p(x,y) = \sum_x p(x,y) = 1 \quad \textit{and} \quad p(x,y) \geq 0$$

Exercise 3. *Let* $\{\mathbf{X}_i\}_i$ *be a stationary Markov chain. Show that* $H(\mathbf{X}_n \mid \mathbf{X}_1)$ *increases with* n *by a direct method and by using the data processing inequality (see Exercise session* 1*).*

Exercise 4 (Second law of thermodynamics). *In statistical thermodynamics, entropy is often defined as the log of the number of microstates in the system. This corresponds exactly to our notion of entropy if all the states are equally likely. But why does entropy increase?*

We model the isolated system as a Markov chain with transitions obeying the physical laws governing the system. Implicitly in this assumption is the notion of an overall state of the system and the fact that knowing the present state, the future of the system is independent of the past. In such a system we can find four different interpretations of the second law. It may come as a shock to find that the entropy does not always increase. However, relative entropy always decreases.

1. **Chain-rule for the Kullback-Leibler divergence.** *Let $p(\cdot)$ and $q(\cdot)$ be two distributions. By definition,*

$$D_{\mathrm{KL}}(p(y \mid x)||q(y \mid x)) = \sum_{x,y} p(x,y) \log_2 \frac{p(y \mid x)}{q(y \mid x)}$$

Show that,

$$D_{\mathrm{KL}}(p(y \mid x)||q(y \mid x)) \geq 0$$

and,

$$D_{\mathrm{KL}}(p(x,y)||q(x,y)) = D_{\mathrm{KL}}(p(x)||q(x)) + D_{\mathrm{KL}}(p(y \mid x)||q(y \mid x))$$

2. **Relative entropy $D_{\mathbf{KL}}(\mu_n||\mu'_n)$ decreases with n.** *Let μ_n and μ'_n be the distributions associated to the same Markov chain of order 1 but with different initial conditions. Show that, $D_{\mathrm{KL}}(\mu_n||\mu'_n)$ decreases with n.*

3. *Let μ be a stationary distribution of the Markov chain. Show that $D_{\mathrm{KL}}(\mu_n||\mu)$ decreases with n.*

4. **Entropy increases if the stationary distribution is uniform**. *Suppose that μ is the uniform distribution. Show that the entropy of μ_n increases.*

Exercise 5 (Shuffles increase entropy)**.** *Let $\mathbf{T}$ be a shuffle (permutation) of a deck of cards and $\mathbf{X}$ be the initial (random) position of the cards in the deck. Our aim is to show that shuffling increases the entropy,*

$$H(\mathbf{TX}) \geq H(\mathbf{X})$$

Justify, informally and formally, each of the following inequalities when $\mathbf{X}$ and $\mathbf{T}$ are independent,

$$\begin{aligned} H(\mathbf{TX}) &\geq H(\mathbf{TX} \mid \mathbf{T}) \\ &= H(\mathbf{T}^{-1}\mathbf{TX} \mid \mathbf{T}) \\ &= H(\mathbf{X} \mid \mathbf{T}) \\ &= H(\mathbf{X}) \end{aligned}$$

Exercise 6 (Entropy rates of Markov chains)**.**

1. *Find the entropy rate of the two-state Markov chain with transition matrix*

$$\mathbf{P} = \begin{pmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{pmatrix}$$

2. *What values of p_{01}, p_{10} maximize the entropy rate?*

3. *Find the entropy rate of the two-state Markov chain with transition matrix*

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}$$

4. *Find the maximum value of the entropy rate of the Markov chain of the previous question (don't forget that $p \leq 1$).*

5. *Let $N(t)$ be the number of allowable state sequences of length t for the Markov chain of question 3. Find $N(t)$ and calculate*

$$H_0 \overset{def}{=} \lim_{t \to +\infty} N(t)$$

Hint: *Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is H_0 an upper bound on the entropy rate of the Markov chain? Compare H_0 with the maximum entropy found in Question 4.*

Exercise 7 (The past has little to say about the future)**.** *Show that for a stationary process $\{\mathbf{X}_i\}_i$,*

$$\frac{1}{n}\, I\Big((\mathbf{X}_1, \ldots, \mathbf{X}_n), (\mathbf{X}_{n+1}, \ldots, \mathbf{X}_{2n})\Big) \underset{n \to +\infty}{\longrightarrow} 0$$

Exercise 8 (Entropy rate of a dog looking for a bone)**.** *A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $\mathbf{X}_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:*

$$(\mathbf{X}_0, \mathbf{X}_1, \ldots,) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots).$$

1. *Compute $H(\mathbf{X}_1, \ldots, \mathbf{X}_n)$*

2. *Find the entropy rate of the dog*

3. *What is the expected number of steps that the dog takes before reversing direction?*

Exercise 9 (Wait before parking your car)**.** *We modelize a car as follows:*

- *if on day n the car is working, it is broken on day $n+1$ with probability p_0*
- *if on day n the car is broken, it is working again on day $n+1$ with probability p_1*

1. *Modelize this process by a Markov chain $\{\mathbf{X}_i\}$ where $\mathbf{X}_i \in \{0,1\}$ denotes the event the car is working or not at day i*
2. *Suppose that $0 < |p_0 p_1| < 1$ and that at day 0 the car is working. Give the asymptotic development of W_N the random variable modelizing the number of days a car is working over a period of N days.*
3. *What does happen if $p_0 = p_1 = 1$? What do you conclude?*