SAMPLE SOLUTIONS EXERCISE 4

EXERCISE 4.1: MAXIMAL ENTROPY (3P)

The Shannon entropy of a probability distribution is given by

$$H = -\sum_c p_c \ln p_c \,.$$

(a) Use Jensens inequality for convex functions to show that the Shannon entropy attains its global maximum for a uniform distribution. (1P)

(b) Prove the same statement in the framework of variational calculus. Use the method of Langrange multipliers to take the normalization constraint of the probability distribution into account. (2P)

SAMPLE SOLUTION

(a) The claim of Jensen's inequality is the following: If a random variable X is mapped by a convex function f, the corresponding expectation values $\langle ... \rangle$ will obey the inequality

$$f\big(\langle X\rangle\big) \leq \langle f(X)\rangle \,.$$

Applied to a probability distribution $\{p_c\}$ this implies the inequality

$$f\Big(\sum_c p_c x_c\Big) \leq \sum_c p_c f(x_c)$$

Let us now choose $f(x_c) = \ln(x_c)$ and $x_c = 1/p_c$. Here one has to take into account that the logarithm is not a convex function, it is rather concave so that we have to flip the inequality. This leads to the following expression for the entropy

$$H = -\sum_c p_c \ln p_c = +\langle \ln(1/p_c)\rangle \leq \ln\Big(\langle 1/p_c\rangle\Big) = \ln \sum_c p_c \frac{1}{p_c} = \ln|\Omega| \,.$$

Therefor, H is bounded from above by $\ln|\Omega|$. In the case of equally probable configurations (microcanonical equilibrium) this value is actually reached.

(b) The aim is to maximize the Shannon entropy $H = -\sum_c p_c \ln p_c$ under the constraint of the normalization $\sum_c p_c = 1$. To this end we apply the method of Lagrange multiplier, minimizing the functional

$$F = -\sum_c p_c \ln p_c + \lambda\Big(\sum_c p_c - 1\Big) \,,$$

where we assume the p_c to be independent. This results into $|\Omega|+1$ equations, namely:

$$0 = -\ln p_c - 1 + \lambda \qquad \forall c \in \Omega$$

$$0 = \sum_c p_c - 1$$

The first equation gives

$$p_c = e^{-1+\lambda} .$$

Therefore, one obtains one solution which is proportional to a constant distribution. The second multiplier has to be chosen in such a way that the second equation is obeyed. This results into a normalized constant distribution of maximal entropy.

Exercise 4.2: Majorization of probability distributions (9P)

Let $\mathbf{a} = \{a_1, a_2, \ldots, a_N\}$ and $\mathbf{b} = \{a_1, a_2, \ldots, a_N\}$ be two sets of numbers, both sorted in descending order. The set $\mathbf{a}$ is said to *majorize* the set $\mathbf{b}$ (denoted as as $\mathbf{a} \succ \mathbf{b}$) if

$$\mathbf{a} \succ \mathbf{b} \qquad \Leftrightarrow \qquad \sum_{i=1}^{n} a_i \geq \sum_{i=1}^{n} b_i \text{ for all } n = 1, \ldots, N-1 \text{ and } \sum_{i=1}^{N} a_i = \sum_{i=1}^{N} b_i$$

(a) Consider a Markov process in a closed system (symmetric rates). Let $\mathbf{p}(t) = \{P_s(t)\}$ be the actual set of probabilities at time t sorted in descending order. Use the master equation to prove that $\mathbf{p}(t) \succ \mathbf{p}(t')$ for $t < t'$. (2P)

(b) For given numbers $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ let

$$\begin{aligned} x_i &:= X_{i+1} - X_i \qquad && \text{for } i = 1, \ldots, N-1 \\ y_i &:= Y_i - Y_{i-1} \qquad && \text{for } i = 2, \ldots, N \\ y_1 &= Y_1 \end{aligned}$$

Prove Abel's partial sum theorem (a discrete version of partial integration): (2P)

$$\sum_{i=1}^{N} X_i y_i \;=\; X_N Y_N - \sum_{i=1}^{N-1} x_i Y_i \, .$$

(c) Let $f(x)$ be a concave function. Prove that $\mathbf{a} \succ \mathbf{b}$ implies that (3P)

$$\sum_{i=1}^{N} f(a_i) \;\leq\; \sum_{i=1}^{N} f(b_i) \, .$$

Hint: Apply (b) to the expression

$$\sum_{i=1}^{N} \Big(f(b_i) - f(a_i) \Big) \;=\; \sum_{i=1}^{N} \Big(\underbrace{\frac{f(b_i) - f(a_i)}{b_i - a_i}}_{=X_i} \underbrace{(b_i - a_i)}_{=y_i} \Big)$$

and try to find an inequality for *secant slopes* on concave functions.

(d) Use (c) to show that for two probability distributions with $\mathbf{p} \succ \mathbf{q}$, the Shannon entropy H satisfies the inequality $H(\mathbf{q}) \geq H(\mathbf{p})$. (1P)

Note: Together with (a), this confirms the Second Law of Thermodynamics.

(e) Show the same for the Rényi entropy H_α. (1P)

Sample Solution

(a) Let $N = |\Omega|$ be the number of microstates and let us assume that the $P_i(t)$ are sorted in descending order with i running from 1 to N. Since probability distributions are normalized by $\sum_{i=1}^N P_i(t) = 1$, the second condition (viz. $\sum_{i=1}^N a_i = \sum_{i=1}^N b_i$) is satisfied automatically. Let us consider the cumulative distribution

$$C_n(t) = \sum_{i=1}^n P_i(t)\,.$$

Now let us compute the temporal derivative and insert the master equation: (1P)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}C_n(t) &= \sum_{i=1}^n \frac{\mathrm{d}}{\mathrm{d}t}P_i(t)\\ &= \sum_{i=1}^n\Big(\sum_{j=1}^N P_j(t)w_{j\to i} - \sum_{j=1}^N P_i(t)w_{i\to j}\Big) \qquad \mid \text{use symmetric rates}\\ &= \sum_{i=1}^n\sum_{j=1}^N \big(P_j(t)-P_i(t)\big)w_{i\to j}\end{aligned}$$

Now the trick is to split the second sum $\sum_{i=1}^N$ into $\sum_{i=1}^n$ and $\sum_{i=n+1}^N$: (1P)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}C_n(t) &= \sum_{i=1}^n\Big[\sum_{j=1}^n\big(P_j(t)-P_i(t)\big)w_{i\to j} + \sum_{j=n+1}^N\big(P_j(t)-P_i(t)\big)w_{i\to j}\Big]\\ &= \underbrace{\sum_{i=1}^n\sum_{j=1}^n\big(P_j(t)-P_i(t)\big)w_{i\to j}}_{=0} + \sum_{i=1}^n\sum_{j=n+1}^N\underbrace{\big(P_j(t)-P_i(t)\big)}_{\leq 0}\underbrace{w_{i\to j}}_{\geq 0} \;\leq\; 0\end{aligned}$$

Therefore $\mathbf{p}(t) \succ \mathbf{p}(t')$ for $t < t'$.

Note: The first term vanishes because of the antisymmetry of the bracket and the symmetry of the rate under exchange of i and j. In the second term the round bracket is negative because of the descending order. Rates are by definition non-negative.

(b) Proof of Abel's partial summation theorem: (2P)

$$\begin{aligned}\sum_{i=1}^N X_i y_i &= X_1 y_1 + \sum_{i=2}^N X_i y_i \;=\; X_1 Y_1 + \sum_{i=2}^N X_i(Y_i - Y_{i-1})\\ &= X_1Y_1 + X_2(Y_2-Y_1) + X_3(Y_3-Y_2) + \ldots + X_N(Y_N - Y_{N-1})\\ &= (X_1-X_2)Y_1 + (X_2-X_3)Y_2 + \ldots + (X_{N-1}-X_N)Y_{N-1} + X_NY_N\\ &= \sum_{i=1}^{N-1}\underbrace{(X_i - X_{i+1})}_{=-x_i}Y_i + X_NY_N \;=\; X_NY_N - \sum_{i=1}^{N-1} x_iY_i\end{aligned}$$

Correction advice: 1P for reorganizing the sum, 1P for the correct final result.

(c) According to the hint we apply Abel's theorem:

$$\sum_{i=1}^{N}\Big(f(b_i)-f(a_i)\Big) \;=\; \sum_{i=1}^{N}\Big[\underbrace{\frac{f(b_i)-f(a_i)}{b_i-a_i}}_{=X_i}\;\underbrace{(b_i-a_i)}_{=y_i}\Big]$$

First we note that

$$X_i \;=\; \sum_{j=1}^{i} y_j \;=\; \sum_{j=1}^{i} a_j - \sum_{j=1}^{i} b_j$$

In particular, according to our assumptions in the exercise, we have (1P)

$$X_N \;=\; \sum_{j=1}^{N} a_j - \sum_{j=1}^{N} b_j \;=\; 0$$

implying that the 'boundary term' $X_N Y_N$ of the 'partial integration' in Abel's formula vanishes. Thus we are left with the expression (1P)

$$\sum_{i=1}^{N}\Big(f(b_i)-f(a_i)\Big) \;=\; -\sum_{i=1}^{N-1}\Big[\Big(\underbrace{\frac{f(b_{i+1})-f(a_{i+1})}{b_{i+1}-a_{i+1}}}_{A}-\underbrace{\frac{f(b_i)-f(a_i)}{b_i-a_i}}_{B}\Big)\underbrace{\Big(\sum_{j=1}^{i} b_j-\sum_{j=1}^{i} a_j\Big)}_{C}\Big]$$

Now the main insight is that A and B are the *secant slops* of the concave function $f(x)$ evaluated between a_i, b_i and $a_{i+1}b_{i+1}$. Since they are sorted in descending order, the secant slope A is evaluated 'left from' B. Since f is concave, this implies that $A \geq B$. Furthermore, knowing that $\mathbf{a} \succ \mathbf{b}$, it is clear that $C \leq 0$. (1P)

Altogether we therefore arrive at the inequality

$$\sum_{i=1}^{N}\Big(f(b_i)-f(a_i)\Big) \;\geq\; 0.$$

(d) In order to study the Shannon entropy $H(\mathbf{p}) = -\sum_{i=1}^{N} p_i \ln p_i$, we consider the function $f(x) = -x\ln x$. Its second derivative is $f''(x) = -1/x$, hence $f(x)$ is strictly concave. If $\mathbf{p} \succ \mathbf{q}$ we have

$$0 \leq \sum_{i=1}^{N}\Big(f(q_i)-f(p_i)\Big)$$

implying that $H(\mathbf{q}) \geq H(\mathbf{p})$.
Note: Since we have shown in (a) that earlier probability distributions in isolated systems majorize later ones, the inequality implies that the Shannon entropy cannot decrease as time proceeds.

(e) The Rényi entropy is defined as

$$H_\alpha(\mathbf{p}) \;=\; \frac{1}{1-\alpha}\ln\Big(\sum_{i=1}^{N} p_i^\alpha\Big).$$

where $\alpha > 0$. Let us consider the term in the bracket, which corresponds to taking $f(x) = x^\alpha$. Since $f''(x) = \alpha(\alpha - 1)x^{\alpha-2}$, this function is convex (concave) for $\alpha > 1$ ($0 < \alpha < 1$). Let us first study the case $0 < \alpha < 1$. If $\mathbf{p} \succ \mathbf{q}$ we get again the inequality

$$\sum_{i=1}^{N} \Big(f(q_i) - f(p_i) \Big) \ = \ \sum_{i=1}^{N} q_i^\alpha - \sum_{i=1}^{N} p_i^\alpha \ \geq \ 0$$

and since the logarithm is monotonously increasing and the prefactor $\frac{1}{1-\alpha}$ is positive, this leads to

$$H_\alpha(\mathbf{q}) \geq H_\alpha(\mathbf{p}).$$

For $\alpha > 1$ the function is convex and the inequality flips from $\geq$ to $\leq$, but on the other hand the prefactor becomes negative, compensating the flip. So we get the same result also in this case. The limit $\alpha \to 1$ is the Shannon limit, hence we consistently recover (d).

Note: We can conclude that the Rńyi entropy respects majorization (= obeys the Second Law) for any value of the paramter $\alpha > 0$. In other words, this gives us *infinitely many* Second Laws, namely, for each value of the parameter α.

(Σ = 12P)