Sample Solutions Exercise 7

Exercise 7.1: Previous knowledge in the form of constraints (6P)

Consider a random variable with three possible values $X \in \{1,2,3\}$. Let $P(1), P(2), P(3)$ be the corresponding discrete probability distribution. The purpose of this exercise is to demonstrate that in the presence of constraints (*Zwangsbedingungen*), the probability distribution is the one which maximizes the Shannon entropy H under these constraints.

(a) Use the method of Lagrange multipliers[1] to find the values of $P(1), P(2), P(3)$ for which the Shannon entropy $H = -\sum_{n=1}^{3} P(n) \ln P(n)$ becomes extremal under the constraint that the probability distribution is normalized: $\sum_{n=1}^{3} P(n) = 1$. Compute the expectation value $\langle x \rangle$ and the variance $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ for this solution. (2P)

(b) Repeat the calculation with the *additional* constraint that the expectation value $\langle x \rangle = \sum_{n=1}^{3} nP(n)$ equals 4/3. Compute the variance σ^2 for this solution. Note that you need two Lagrange multipliers in this case. (2P)

(c) Determine the normalized probability distribution under the constraints $\langle x \rangle = 4/3$ and a fixed given variance σ^2. Determine the range in which σ^2 can be chosen so that the solution is a valid probability distribution. (2P)

You may use Mathematica® or similar algebraic computer systems to solve the systems of equations in this exercise.

Sample Solution

(a) Let $N = \sum_{x=1}^{3} P(x)$. Then the normalization constraint reads $N = 1$ or $N - 1 = 0$. Therefore, we define the Lagrangian function

$$L = H + \lambda(N-1) = \Big(-\sum_x P(x) \ln P(x)\Big) + \lambda\Big(\sum_x P(x) - 1\Big),$$

where λ is the Lagrange multiplier associated with the normalization. Then the four equations to be solved read:

$$\frac{\partial L}{\partial P(1)} = -1 + \lambda - \ln P(1) = 0\,,$$
$$\frac{\partial L}{\partial P(2)} = -1 + \lambda - \ln P(2) = 0\,,$$
$$\frac{\partial L}{\partial P(3)} = -1 + \lambda - \ln P(3) = 0\,,$$
$$\frac{\partial L}{\partial \lambda} = P(1) + P(2) + P(3) - 1 = 0\,.$$

The first three equations imply that $P(1) = P(2) = P(3)$, that is, the probability distribution is uniform. The last one fixes the norm $P(1) = P(2) = P(3) = \frac{1}{3}$. This gives the mean value $\langle x \rangle = 2$ and the variance $\sigma^2 = \frac{2}{3}$.

[1] see e.g. https://de.wikipedia.org/wiki/Lagrange-Multiplikator

(b) Now we add the additional constraint that $\langle x\rangle = \frac{4}{3}$, i.e., $\sum_x xP(x) - \frac{4}{3} = 0$. The Lagrangian function now reads

$$\begin{aligned} L &= H + \lambda(N-1) + \mu(\langle x\rangle - \frac{4}{3}) \\ &= \Big(-\sum_x P(x)\ln P(x)\Big) + \lambda\Big(\sum_x P(x) - 1\Big) + \mu\Big(\sum_x xP(x) - \frac{4}{3}\Big), \end{aligned}$$

where μ is another Lagrange multiplier connected with the constraint on the expectation value. Now we have a system of five equation, namely

$$\begin{aligned} \frac{\partial L}{\partial P(1)} &= -1 + \lambda + \mu - \ln P(1) = 0\,, \\ \frac{\partial L}{\partial P(2)} &= -1 + \lambda + 2\mu - \ln P(2) = 0\,, \\ \frac{\partial L}{\partial P(3)} &= -1 + \lambda + 3\mu - \ln P(3) = 0\,, \\ \frac{\partial L}{\partial \lambda} &= P(1) + P(2) + P(3) - 1 = 0\,. \\ \frac{\partial L}{\partial \mu} &= P(1) + 2P(2) + 3P(3) - \frac{4}{3} = 0\,. \end{aligned}$$

Using e.g. *Mathematica®* we find the solution

$$P(1) = \frac{1}{9}(9-\sqrt{6})\,, \quad P(2) = \frac{1}{9}(-3+2\sqrt{6})\,, \quad P(3) = \frac{1}{9}(3-\sqrt{6})$$

with

$$\lambda = 1 - \ln\Big[\frac{1}{9}(3+8\sqrt{6})\Big]\,, \quad \mu = \ln\Big[\frac{1}{5}(-1+\sqrt{6})\Big]\,.$$

The corresponding numerical values are

$P(1) = 0.727834$, $P(2) = 0.210998$, $P(3) = 0.0611678$, $\lambda = 1.92054$, $\mu = -1.23823$.

This allows us to compute the expectation value

$$\langle x\rangle = \frac{4}{3}\,,$$

which of course reproduces the constraint, and the variance

$$\sigma^2 = \langle x^2\rangle - \langle x\rangle^2 = \frac{2}{9}(\sqrt{6}-4) \approx 0.344558.$$

Note: The new constraint reduces the variance. This is intuitively clear: Without constraining the expectation value, the system has more freedom to fluctuate.

(c) Fixing norm, expectation value and variance we have already three equation. In this case we do not need the Lagrange technique. (We could, of course, add another Lagrange multiplier, and this would also lead to the correct result). So we have to solve the system of equations

$$\sum_x P(x) = 1\,, \quad \sum_x xP(x) = \frac{4}{3}\,, \quad \sum_x x^2P(x) - \frac{16}{9} = \sigma^2.$$

Solving this system with an algebraic computer system yields the probability distribution

$$P(1) = \frac{1}{18}(10 + 9\sigma^2)\,, \quad P(2) = \frac{1}{9}(5 - 9\sigma^2)\,, \quad P(3) = \frac{1}{18}(9\sigma^2 - 2).$$

The probability distribution is valid as long as these probabilities are real-valued number between 0 and 1. This gives rise to three inequalities of the form

- For $P(1)$: $-\frac{10}{9} \le \sigma^2 \le \frac{8}{9}$
- For $P(2)$: $-\frac{4}{9} \le \sigma^2 \le \frac{5}{9}$
- For $P(3)$: $+\frac{2}{9} \le \sigma^2 \le \frac{20}{9}$

Intersecting them gives:

$$\frac{2}{9} \le \sigma^2 \le \frac{5}{9}$$

As a cross-check, we note that the value of σ^2 obtained in (b) is in this range.

Exercise 7.2: Random switch **(6P)**

Consider a switch which is either on (1) or off (0). Assume that the switch is initially off at $t = 0$ and changes its state randomly with the rates $w_{0\to 1}$ and $w_{1\to 0}$.

(a) Write down the master equation and compute $P_0(t)$ and $P_1(t)$. (2P)

(b) Compute the Shannon entropy $H(t)$ for symmetric rates $w_{0\to 1} = w_{1\to 0} = 1$. Check the Second Law by plotting $H(t)$ in the range $t = 0...2$. (2P)

(c) For arbitrary rates the Second Law does not apply. Find a condition for which the entropy first increases, then reaches a maximum at a finite time t_{max} which is followed by a decrease. Compute t_{max} and $H(t_{max})$ as well as the asymptotic saturation value $\lim_{t\to\infty} H(t)$. (2P)

Sample Solution

(a) We have $P_0(t) + P_1(t) = 1$ for all t. The Master equation reads (1P)

$$\frac{\mathrm{d}}{\mathrm{d}t} P_0(t) \;=\; w_{1\to 0}\big(1 - P_0(t)\big) - w_{0\to 1} P_0(t).$$

This differential equation is trivial to solve. The solution is essentially given by an exponential function. More specifically, with the initial condition $P_0(0) = 1$ we get

$$P_0(t) = \frac{w_{1\to 0} + w_{0\to 1} e^{-t(w_{1\to 0} + w_{0\to 1})}}{w_{1\to 0} + w_{0\to 1}}$$

and $P_1(t) = 1 - P_0(t)$. (1P)

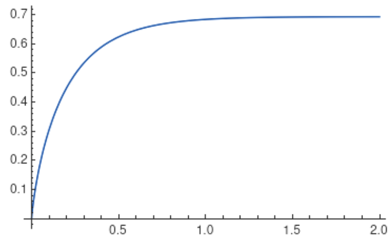
(b) For symmetric rates $w_{0\to1} = w_{1\to0} = 1$ the above result simplifies to

$$P_0(t) \;=\; \frac{1}{2}\left(1 + e^{-2t}\right).$$

The corresponding Shannon entropy reads: (1P)

$$H(t) = \frac{1}{2}e^{-2t}\left(-\left(e^{2t}-1\right)\ln\left(\frac{1}{2}-\frac{e^{-2t}}{2}\right)-\left(e^{2t}+1\right)\ln\left(\frac{1}{2}\left(e^{-2t}+1\right)\right)\right)$$

This is how it looks like. As can be seen, it increases monotonically, in accordance with the Second Law, until it reaches the uncertainty of $ln2 = 1bit$: (1P)



(c) For arbitrary rates the entropy is a somewhat more complicated function:

$$H(t) = \frac{e^{-t(\mathrm{w01}+\mathrm{w10})}\left(\mathrm{w01}\left(-\left(e^{t(\mathrm{w01}+\mathrm{w10})}-1\right)\right)\log\left(\frac{\mathrm{w01}-\mathrm{w01}e^{-t(\mathrm{w01}+\mathrm{w10})}}{\mathrm{w01}+\mathrm{w10}}\right)-\left(\mathrm{w10}e^{t(\mathrm{w01}+\mathrm{w10})}+\mathrm{w01}\right)\log\left(\frac{\mathrm{w01}e^{-t(\mathrm{w01}+\mathrm{w10})}+\mathrm{w10}}{\mathrm{w01}+\mathrm{w10}}\right)\right)}{\mathrm{w01}+\mathrm{w10}}$$

We can use *Mathematica*® for finding an extremum (setting the derivative to zero). The solution reads (0.5P)

$$t_{max} = \frac{1}{w_{0\to1}+w_{1\to0}}\ln\left(\frac{2w_{0\to1}}{w_{0\to1}-w_{1\to0}}\right)$$

Since the argument of the logarithm has to be positive, we can conclude that such an intermediate maximum exists if (0.5P)

$$\boxed{w_{0\to1} > w_{1\to0}.}$$

Note that this result depends, of course, on the choice of the initial condition made above. If this condition holds, the entropy in the maximum turns out to be (0.5P)

$$H(t_{max}) = \ln 2$$

corresponding to 1 *bit* uncertainty. For $t > t_{max}$ the entropy decreases again, saturating at (0.5P)

$$H(\infty) \;=\; \lim_{t\to\infty} H(t) \;=\; \frac{w_{1\to0}}{w_{0\to1}+w_{1\to0}}.$$

(Σ = 12P)