3. ... and Back Again: The Maximum Information Principle (MIP)

3.1 Some Basic Ideas

In this chapter we address the following question: Let some macroscopic quantities of a system be given. We then wish to devise a procedure by which we can derive the probability distribution of macroscopic or even microscopic variables. In other words, we start from the macroscopic world and wish to draw conclusions about the microscopic world. Depending on the kind of systems we are treating, the adequate macroscopic quantities may be quite different. In closed physical systems, to which thermodynamics applies, these quantities are energy, particle numbers etc., and we shall illustrate the general procedure by this example in Chap. 4. In open systems, e.g. in physics or biology, the adequate macroscopic quantities will turn out to be, for instance, intensities and intensity fluctuations. Indeed, it will be the main purpose of the following chapters, to deal with open systems.

Since the starting point of our approach is the concept of information, we shall derive this concept in this Sect. 3.1.

By some sort of new interpretation of probability theory we get an insight into a seemingly quite different discipline, namely information theory. Consider the sequence of tossing a coin with outcomes 0 and 1. Now interpret 0 and 1 as a dash and dot of a Morse alphabet. We all know that by means of a Morse alphabet we can transmit messages so that we may ascribe a certain meaning to a certain sequence of symbols. In other words, a certain sequence of symbols carries information. In information theory we try to find a measure for the amount of information.

Let us consider a simple example and consider \( R_0 \) different possible events ("realizations") which are equally probable a priori. Thus when tossing a coin we have the events 1 and 0 and \( R_0 = 2 \). In the case of a die we have 6 different outcomes, therefore \( R_0 = 6 \). Thus the outcome of tossing a coin or throwing a die is interpreted as the receipt of a message and only one out of the possible \( R_0 \) outcomes is actually realized. Apparently the greater \( R_0 \), the greater is the uncertainty before the message is received and the larger will be the amount of information after the message is received. Thus we may interpret the whole procedure in the following manner: In the initial situation we have no information \( I_0 \), i.e., \( I_0 = 0 \) with \( R_0 \) equally probable outcomes.

In the final situation we have an information \( I_1 \neq 0 \) with \( R_1 = 1 \), i.e., a single outcome. We now want to introduce a measure for the amount of information, \( I \), which apparently must be connected with \( R_0 \). To get an idea how the connection between \( R_0 \) and \( I \) must appear we require that \( I \) is additive for independent events.
Thus when we have two such sets with \( R_0 \) or \( R_2 \) outcomes so that the total number of outcomes is

\[
R = R_0 R_2
\]  
(3.1)

we require

\[
I(R_0 R_2) = I(R_0) + I(R_2) .
\]  
(3.2)

This relation can be fulfilled by choosing

\[
I = K \ln R_0
\]  
(3.3)

where \( K \) is a constant. It can even be shown that (3.3) is the only solution to (3.2). The constant \( K \) is still arbitrary and can be fixed by some definition. Usually the following definition is used. We consider a so-called “binary” system which has only two symbols (or letters). These may be the head and the tail of a coin, or answers yes and no, or numbers 0 and 1 in a binomial system. When we form all possible “words” (or sequences) of length \( n \), we find \( R = 2^n \) realizations. We now want to identify \( I \) with \( n \) in such a binary system. We therefore require

\[
I = K \ln R \equiv K \ln 2 = n
\]  
(3.4)

which is fulfilled by

\[
K = \frac{1}{\ln 2} = \log_2 e
\]  
(3.5)

With this choice of \( K \), another form of (3.4) reads

\[
I = \log_2 R .
\]  
(3.4a)

Since a single position in a sequence of symbols (signs) in a binary system is called “bit”, the information \( I \) is now directly given in bits. Thus if \( R = 8 = 2^3 \) we find \( I = 3 \) bits and generally for \( R = 2^n \), \( I = n \) bits. The definition of information for (3.3) can be easily generalized to the case where we initially have \( R_0 \) equally probable cases and finally \( R_1 \) equally probable cases. In this case the information is

\[
I = K \ln R_0 - K \ln R_1
\]  
(3.6)

which reduces to the earlier definition (3.3), if \( R_1 = 1 \). A simple example for this is given by a die. Let us define a game in which the even numbers mean gain and the odd numbers mean loss. Then \( R_0 = 6 \) and \( R_1 = 3 \). In this case the information content is the same as that of a coin with originally just two possibilities.

We now derive a more convenient expression for the information: We first consider the following example of a simplified Morse alphabet with dash and dot (in the real Morse alphabet, the intermission is a third symbol). We consider a word of length \( G \) which contains \( N_1 \) dashes and \( N_2 \) dots, with

\[
N_1 + N_2 = N .
\]  
(3.7)