2.4 The Entropy

An important characteristic feature for a random variable is the entropy, which we will introduce in this section.

2.4.1 Entropy for a Discrete Set of Events

Let $\{A_1, \ldots, A_N\}$ be a complete, disjoint set of events, i.e.,

$$A_1 \cup A_2 \cup \ldots \cup A_N = \Omega \,. \tag{2.78}$$

Furthermore, let $\mathcal{P}$ be a probability defined for these events. We then define the entropy as

$$S = -k \sum_{i=1}^{N} \mathcal{P}(A_i) \ln(\mathcal{P}(A_i)) \,. \tag{2.79}$$

Here k represents a factor which we set equal to 1 for the moment. In the framework of statistical mechanics k will be Boltzmann's constant k_B.

We observe:

- The entropy is defined for a complete, disjoint set of events of a random variable, irrespective of whether this partition of Ω into events can be refined or not. If Ω is the real axis, we might have, e.g., $N = 2, A_1 = (-\infty, 0), A_2 = [0, \infty)$.
- Since $0 \le \mathcal{P}(A_i) \le 1$ we always have $S \ge 0$.
- If $\mathcal{P}(A_j) = 1$ for a certain j and $\mathcal{P}(A_i) = 0$ otherwise, then $S = 0$. This means that if the event A_j occurs with certainty the entropy is zero.
- If an event has occurred, then, as we will show in a moment, $-\log_2 \mathcal{P}(A_j)$ is a good measure of the number of questions to be asked in order to find out that it is just A_j which is realized. In this context, 'question' refers to questions which can be answered by 'yes' or 'no', i.e., the answer leads to a gain of information of 1 bit. Hence, on average the required number of yes-or-no questions is

$$S' = -\sum_{j=1}^{N} \mathcal{P}(A_j) \log_2(\mathcal{P}(A_j)) = S + \text{const} \,. \tag{2.80}$$

The entropy is thus a measure of the missing information needed to find out which result is realized.

To show that $-\log_2 \mathcal{P}(A_j)$ is just equal to the number of required yes-or-no questions, we first divide Ω into two disjoint domains Ω_1 and Ω_2 such that

$$\sum_{A_i \in \Omega_1} \mathcal{P}(A_i) = \sum_{A_i \in \Omega_2} \mathcal{P}(A_i) = \frac{1}{2} \,. \tag{2.81}$$

The first question is now: Is A_j in Ω_1? Having the answer to this question we next consider the set containing A_j and multiply the probabilities for the events in this set by a factor of 2. The sum of the probabilities for this set is now again equal to 1, and we are in the same position as before with the set Ω: We divide it again and ask the corresponding yes-or-no question. This procedure ends after k steps, where k is the smallest integer such that $2^k\,\mathcal{P}(A_j)$ becomes equal to or larger than 1. Consequently, $-\log_2 \mathcal{P}(A_j)$ is a good measure of the number of yes-or-no questions needed.

• If the probabilities of the events are equal, i.e.,

$$\mathcal{P}(A_i) = \frac{1}{N}\,, \tag{2.82}$$

we have

$$S = \ln N\,. \tag{2.83}$$

Any other distribution of probabilities leads to a smaller S. This will be shown soon.

The above observations suggest that the entropy may be considered as a lack of information when a probability density is given. On average it would require the answers to S yes-or-no questions to figure out which event has occurred. This lack is zero for a density which describes the situation where one event occurs with certainty. If all events are equally probable, this lack of information about which event will occur in a realization is maximal.

A less subjective interpretation of entropy arises when we think of it as a measure for uncertainty. If the probability is the same for all events, the uncertainty is maximal.

2.4.2 Entropy for a Continuous Space of Events

In a similar manner we define the entropy for a random variable X, where the space of events is a continuum, by

$$S[\varrho_X] = -k \int \mathrm{d}x\, \varrho_X(x)\, \ln\left(\frac{\varrho_X(x)}{\varrho_0}\right)\,. \tag{2.84}$$

When $\varrho_X(x)$ has a physical dimension, the denominator ϱ_0 in the argument of the logarithm cannot simply be set to 1. Since the physical dimension of $\varrho_X(x)$ is equal to the dimension of $1/\mathrm{d}x$, the physical dimension of ϱ_0 has to be the same, in order that the argument of the logarithm will be dimensionless.

It is easy to see that a change of ϱ_0 by a factor α leads to a change of the entropy by an additive term $k\ln\alpha$. Such a change of ϱ_0 only shifts the scale of S. Notice that we no longer have $S \geq 0$.