Introduction to Statistical Physics

Lecture notes for the Master course

“Computational Science”,

year 2008/09

Johann Wolfgang Goethe University , Frankfurt am Main

lecturer: Christian Holm

14 October 2008

FIAS, Frankfurt Institute for Advanced Studies,
Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

General remarks

- Lectures: Tuesday 13:00-15:00, FIAS room 100.
- Tutorials: Tuesday 15:00-17:00, FIAS room 100.
- The attendance to lectures and tutorials is mandatory. If the student fails to show up to a lecture of tutorial, she/he must properly justify it.
- In each tutorial a ”Problem Sheet” will be hand out to the students. The exercises in the ”problem Sheet” must be done individually by each student and returned with the solutions in the following tutorial session. Each ”Problem Sheet” will be evaluated on a base of 10 points distributed among the different problems. Failure to return the ”problem Sheet” on due date will imply 0 points for that ”problem Sheet”. No grace periods will be allowed. At the begginning of each tutorial, a short test will be passed about previous lessons, students can score up to 2 points with each test
- To pass the subject, the student is required to obtain at least a 60% of the total amount of points from tutorials.
- A recommended book to follow these lectures is:

```
Thermal Physics
C.Kittel and H. Kroemer
W.H. Freeman and Company, New York, 1980.
ISBN: 0-7167-1088-9

Niederursel Library: 1 English / 3 German
Shelf mark:           Ta42 (4)
```

- See further information in *http://fias.uni-frankfurt.de/ simbio/Teaching*.
- Person in charge: Dr. Christian Holm, FIAS, Office 2.301, phone $(0)69/798-47505$, fax $(0)69/798-47611$, e-mail: *holm@fias.uni-frankfurt.de*

-

Contents

1 Motivation

1.1 The big picture

The two key questions of modern science are the following:

1. What is the *microscopic* structure of matter? *(High energy physics, the physics of elementary particles, analytical chemistry, biochemistry, genetics)*

2. How do properties of *macroscopic* systems follow from their *many microscopic* constituents? *(Statistical Physics, [analytical chemistry, cell biology])*

The aim of statistical physics is:

> Explain macroscopic phenomena from microscopic laws!
>
> (Maybe also starting from a mesoscopic description)

Examples:

- The meaning of *temperature*, *heat*, *entropy*
- *Equations of state* for gases, liquids, solids… (→ thermostatics)
- *transport* phenomena, *irreversible* processes (→ thermodynamics)

Recent successes:

- How come that water boils at 100° C? Why is there a gaseous, liquid, and a solid phase in the first place? Why does a magnet lose its magnetisation if one heats it up? → *Theory of phase transitions*.
- Why does liquid ^{4}He lose its viscosity below 2.2 K? → *Theory of suprafluidity*.
- Why does copper conduct an electrical current but glass doesn't? → *Theory of electrical conductivity*.
- Where do the structures come from which spontaneously form in microemulsions? → *Theory of pattern formation*.

Methodology of statistical physics

- Appropriate treatment of *nonlinearities* and *large numbers*
- Is important for almost all branches of physics

2 Some basics of probability theory

The notion of "probability" is

- very close to our intuition, but
- quite subtle to formalize mathematically.

For several hundred years mathematicians have been working on how to calculate with probabilities, e.g., Blaise Pascal (1623–1662), Jacob Bernoulli (1654–1705), Johann(III) Bernoulli (1744-1807), Pierre-Simon Laplace (1749–1827) and others. The modern axiomatic treatment goes back to Andrey Nikolaevich Kolmogorov (1903–1987).

We will *not* present probability theory in all its formal glory, but rather try to summarize some essential terminology and techniques necessary for "everyday work". The following good book may be consulted, though:

> William Feller, *An introduction to Probability Theory and Its Application*, Volume I, 3^{rd} edition, John Wiley & Sons, New York (1970).

2.1 Basic notions

We first start with *discrete probability*.

A **random experiment** is an experiment whose outcome depends on *chance*. Every *single outcome* is described by one and only one *point* in an abstract **sample space** $\mathscr{S}$. Any *collection of sample points* is called an **event**. Events are thus *subsets* of the sample space. The points of the sample space are sometimes called "elementary events". Furthermore, to every single point E of some abstract sample space $\mathscr{S}$ belongs a number $P(E)$, which we call the **probability** of E. These numbers satisfy the *normalization condition*

$$\sum_{E\in\mathscr{S}} P(E) = 1 \tag{2.1}$$

A sample space equipped with a probability function is called a **probability space**.[1]

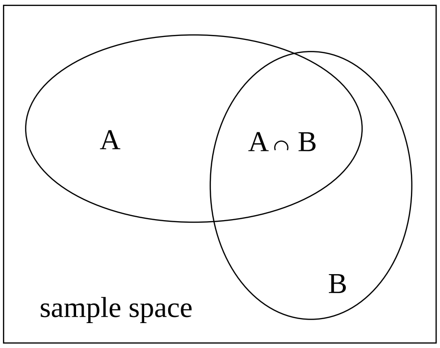
Probability theory thus boils down to calculating with sets, which have some numbers (polite: *measures*) assigned to it. This leads to the two sub-fields of mathematical stochastic:

probability theory How to *calculate* with probabilities

[1] The technical definition is more complicated in continuous spaces, where we need to think a bit more about what it means to assign a measure to a subset.

Statistics How to *determine* the *a priori* probabilities of the elementary events.

Remarks:



Calculating with events is the same as calculating with sets. If A and B are events, it thus makes sense to talk about $A \cup B$ (A or B have occurred), $A \cap B$ (A and B have occurred). If $A \cap B = \emptyset$, the two events A and B are *mutually exclusive*. Elementary events, i.e. points in sample space, are always mutually exclusive. The event $A - B$ contains all points which are in A but not in B. $(A \cup B) - (A \cap B)$ then means the event "A or B but not both at the same time".

There are two special events: $\emptyset$, the *impossible event*, and $\mathscr{S}$, the *certain event*. The event $\overline{A} := \mathscr{S} - A$ is the complementary event of A, i.e., it occurs if and only if A does not occur. It is easy to check that $\overline{A \cap B} = \overline{A} \cup \overline{B}$ and $\overline{A \cup B} = \overline{A} \cap \overline{B}$

The probability A of an event is the sum of the probabilities of all sample points in it.

Hence, the probability of the event $A \cup B$ is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.2}$$

Careful: Only if A and B are mutually exclusive, the last term vanishes, and *only then* then the probability of the union $A \cup B$ is the sum of the probabilities of the individual events!

Two events are called **(stochastically) independent**, if we have $P(A \cap B) = P(A) \cdot P(B)$.

Example

Let's take as our sample space the set of *ordered pairs* (d_1, d_2), where $d_i \in \{1, 2, 3, 4, 5, 6\}$. There are 36 such pairs, and let us say that they all have the same probability, i.e., each pair has the probability $\frac{1}{36}$.

An experimental realization of such a probability space is the throwing of two dice, one red and one blue, so that we can distinguish the first and the second entry in our pair (say, the first belongs to the red die).

The event $A :=$ "the red die showed 2" is realized by the six pairs $(2, 1)$, $(2, 2)$, $(2, 3)$, $(2, 4)$, $(2, 5)$, and $(2, 6)$, and has thus probability $\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$. The event $B :=$ "the sum of the outcomes is even" is realized by the 18 pairs $(1, 1)$, $(1, 3)$, $(1, 5)$, $(2, 2)$, $(2, 4)$, $(2, 6)$, $(3, 1)$, $(3, 3)$, $(3, 5)$, $(4, 2)$, $(4, 4)$, $(4, 6)$, $(5, 1)$, $(5, 3)$, $(5, 5)$, $(6, 2)$, $(6, 4)$, and $(6, 6)$, and its probability is $18 \cdot \frac{1}{36} = \frac{1}{2}$. The event $C :=$ "to throw a double" is realized by the six pairs $(1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, $(5, 5)$, and $(6, 6)$ and also has probability $\frac{1}{6}$. Note that C is a subset of B. Hence, the probability of $B \cap C$ is equal to the probability of B, i.e., $\frac{1}{2}$. On the other hand, $P(B) \cdot P(C) = \frac{1}{12}$. Hence, B and C are not stochastically independent.

2.2 Combinatorics

Calculation of probabilities from elementary probabilities has a lot to do with cleverly *counting all the possibilities which lead to some particular event*. Combinatorics provides a few convenient notions for doing this.

Let us just summarize a few often needed formulas:

(ordered) multiplets We have an ordered multiplet of k elements. At the first position we have n_1 possibilities for some entry, at the second we have n_2, etc. The total number of possibilities to form a multiplet are $n_1 \cdot n_1 \cdots n_k$. If the number of possibilities at each entry is the same, say n, then the total number of possibilities is n^k.

ordered samples Think of a "population" of n symbols $a_1, a_2, \ldots, a_n$. Any ordered arrangement $a_{j_1}, a_{j_2}, \ldots, a_{j_k}$ of k symbols is called an *ordered sample of size* k. One can imagine the symbols being picked one at a time, and one then sees that there exist two different procedures: *with replacement*, i.e., a symbol previously chosen can be chosen again, and *without replacement*, i.e., each symbol can be used only once. In the latter case the sample size cannot exceed the number of available symbols. In the case with replacement the number of different ordered k-samples is of course n^k. But if repetition is not allowed, we have n choices at the first position, but only $n-1$ choices at the second position, $n-2$ choices at the third position and so forth. The number of different k samples is then given by

$$(n)_k := n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) . \tag{2.3}$$

An important special case is if the size of the sample is equal to the number of elements. This for instance shows up in the question "In how many different ways can I order n different elements in a row? The answer is

$$n! := (n)_n = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 . \tag{2.4}$$

Example: What is the probability of having no repetition when picking a sample of size k from a population of n elements? Answer: There are n^k samples in total, of which $(n)_k$ satisfy the no-repetition-condition. If all samples are equally probably, the probability of no repetition is $(n)_k/n^k$. Application: What is the probability of n different persons having their birthday all on different days?

subpopulations Those are the same as samples without repetition, *except that we do not pay attention to order*. So, if two ordered samples have the same elements but in a different order, they would be considered as identical subpopulations. Hence, neglecting order reduces the number of possibilities, but by how much? Take a subpopulation of k elements out of n possibilities. Any arbitrary numbering (ordering) changes it into an ordered sample, and we have $(n)_k$ of them. Since there are $k!$ different ways to number (order) an ordered sample, it follows that there are $k!$ more ordered samples than subpopulations. Hence, there are $(n)_k/k!$ subpopulations, which is usually abbreviated as

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n(n-1)(n-1)\cdots(n-k+1)}{1\cdot 2\cdots(k-1)k} = \frac{n!}{k!(n-k)!} \tag{2.5}$$

Hence, a population of n elements possesses $\binom{n}{k}$ different subpopulations of size $k \leq n$. In other words, a subset of k elements can be chosen in $\binom{n}{k}$ different ways.

Excursion: Stirling's formula

In combinatorics one often has to calculate factorials. This is "hard" because it involves a product over many *different* factors. However, there is a good approximation, originally due to James Stirling (1692–1770), which is often *extremely useful*

$$n! \sim \sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \; . \tag{2.6}$$

By "$\sim$" we mean that the ratio of both sides converges to 1 in the limit $n \to \infty$. Not as accurate, but almost as useful, is the Stirling approximation in the following form:

$$\log n! \sim n \log n - n + 1 \; . \tag{2.7}$$

The proof of Eqn. (2.6) is a bit tricky[2], but can be found for instance in Feller's book. We will show a very sketchy "proof" of the simpler version (2.7). We have

$$\begin{aligned} \log n! &= \log(1\cdot 2\cdot 3\cdots n) \\ &= \log(1) + \log(2) + \log(3) + \cdots + \log(n) \\ &= \sum_{k=1}^{n} \log(k) \\ &\approx \int_1^n \mathrm{d}k \;\log(k) \\ &= n\log n - n + 1 \; . \end{aligned} \tag{2.8}$$

The accuracy of both formulas is illustrated in Fig. 2.1.

2.3 ⋆Conditional probability

Suppose we want to know the probability of a particular event A, but for some reason we already know that the event B occurred also. Our knowledge of B's occurrence will "change" the probability of A, because in some sense we now refer to a different sample space. The probability of A *under the condition that* B *occurred* is usually termed $P(A|B)$ and is called a *conditional probability*.

Example: Look again at the sample space for throwing two dice, as we have used it above – ordered tuples of the numbers $1, \ldots, 6$. Let A be the event that the sum of the two dice is larger than 7, and let B be the event that at least one of the dice shows a '5'.

[2]It is probably most quickly derived as a Laplace-evaluation of the Gamma-function.

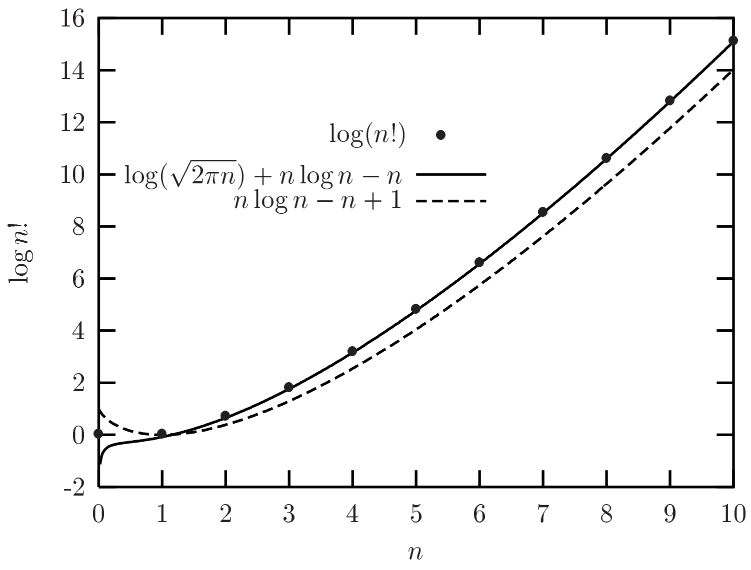


Figure 2.1: Illustration of the accuracy of Stirling's formula. The dots are the exact values of $\log n!$, the solid line is the full formula (2.6), the dashed line is the simpler formula (2.7).

The event of having a sum greater than 7 is realized by the 15 tuples $(2, 6)$, $(3, 5)$, $(3, 6)$, $(4, 4)$, $(4, 5)$, $(4, 6)$, $(5, 3)$, $(5, 4)$, $(5, 5)$, $(5, 6)$, $(6, 2)$, $(6, 3)$, $(6, 4)$, $(6, 5)$, and $(6, 6)$, and thus has probability $P(A) = \frac{15}{36} \approx 0.417$. The event of throwing a 5 with at least one of the dice is realized by the 11 events $(1, 5)$, $(2, 5)$, $(3, 5)$, $(4, 5)$, $(5, 5)$, $(6, 5)$, $(5, 1)$, $(5, 2)$, $(5, 3)$, $(5, 4)$, and $(5, 6)$, so its probability is $P(B) = \frac{11}{36} \approx 0.306$. However, if we already *know* that one of these 11 events occurred, the probability of *now* finding a sum greater than 7 is realized by the 7 tuples $(3, 5)$, $(4, 5)$, $(5, 5)$, $(6, 5)$, $(5, 3)$, $(5, 4)$, and $(5, 6)$. Hence, the probability of finding a sum greater than 7 *given* that we threw at least one 5 is given by $P(A|B) = \frac{7}{11} \approx 0.636$, which in this case is larger than $P(A)$.

It is easy to convince ourselves that the conditional probability can be calculated as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} . \tag{2.9}$$

Of course, this equation does not make sense in the case where $P(B) = 0$. (It does not make sense to ask "what is the probability of occurrence of the event A, given that we know that the impossible event occurred!")

The conditional probabilities of various events with respect to a particular hypothesis H amounts to choosing H as the new sample space with probabilities proportional to the original ones. The proportionality factor $1/P(H)$ is necessary to normalize the new probability space to 1." This shows that *all general results valid for probabilities also hold for conditional probabilities with respect to any hypothesis* H.

Equation (2.9) can also be written as

$$P(A \cap B) = P(A|B)P(B) , \tag{2.10}$$

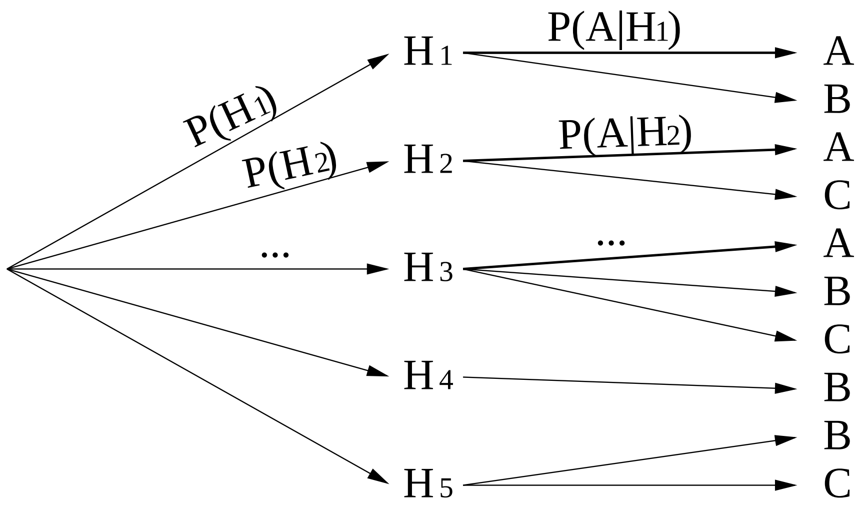


Figure 2.2: Illustration of the calculation of a compound probability by going along mutually exclusive paths.

which is sometimes also called the *theorem of compound probabilities*. Since $P(A \cap B) = P(B \cap A)$, we also see that $P(A|B)P(B)$ must be equal to $P(B|A)P(A)$, from which we find

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \,, \tag{2.11}$$

which shows us how to "invert" condition and event.

Moreover, let $\{H_i\}$ be a set of n mutually exclusive events whose union is the entire sample space, i.e., $H_1 \cup H_2 \cup \ldots \cup H_n = \mathscr{S}$. The we have

$$A = A \cap (H_1 \cup H_2 \cup \ldots \cup H_n) = (A \cap H_1) \cup (A \cap H_2) \cup \ldots \cup (A \cap H_n)\,. \tag{2.12}$$

Since the $A \cap H_i$ are mutually exclusive, their probabilities add. Using Eqn. (2.10) we then find

$$P(A) = \sum_{i=1}^{n} P(A|H_i)P(H_i)\,. \tag{2.13}$$

This formula is useful because an evaluation of the conditional probabilities $P(A|H_i)$ is frequently easier than the direct evaluation of $P(A)$. The formula is graphically illustrated in Fig. 2.2.

Finally, combining Eqn. (2.13) with Eqn. (2.11) yields

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_j P(A|H_j)P(H_j)}\,. \tag{2.14}$$

This is called the *formula of Bayes* or *Bayes' rule for the probability of causes*. In fact, it doesn't really say much more than Eqn. (2.9).

We introduced stochastic independence of two events A and B by requiring that $P(A \cap B) = P(A) \cdot P(B)$. Inserting this into Eqn. (2.10) and canceling $P(B)$ (which hopefully is nonzero), we see that we have

$$P(A) = P(A|B) \qquad \text{(for stochastically independent events)} . \tag{2.15}$$

Hence, whether or not we have knowledge of the occurrence of the event B does not influence the probability of A. This explains why the terminology "stochastic independence".

2.4 Random variables

A random variable is **a function defined on a sample space**. Given any outcome (i.e., elementary event) of a random experiment, the random variable is a function which assigns a value to this outcome.

Example: Take the sample space as the set of *ordered pairs* (d_1, d_2), where $d_i \in \{1, 2, 3, 4, 5, 6\}$ (e.g., throwing of two dice of different color). We could define the random variable S as the function which returns the sum of the numbers which the dice show. Hence, the pair $(2, 3)$ will be mapped by S to the value 5.

Given some random variable X, we can use it to define particular events related to it. Most useful is going to be the following consideration: Given some specific value x, which elementary events will be mapped by the random variable onto this particular value x?

Example: Take the random variable S defined above. Which events are mapped by S onto the value 4? Answer: $(1, 3)$, $(2, 2)$, and $(3, 1)$. Hence, since there are 36 elementary events, the probability of finding an event which is mapped onto the value 4 is $3/36 = 1/12$. In more colloquial terms: The probability of getting the sum 4 when throwing two dice is $1/12$!

We have thus seen that the probability of a random variable having a particular value is quite useful. We will therefore introduce the following definition:

> The probability that an event, which is characterized by the fact that the random variable X takes on the particular value x, is called the **probability distribution of** X. We will write this as
>
> $$w_X(x) \;=\; P(X = x) . \tag{2.16}$$

Example: For the random variable S defined above, the probability distribution $w_S(x)$ looks like this. It is zero everywhere, except for the case when $x \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

And there it has the values

$$w_S : \begin{cases} 2 \longrightarrow 1/36 \approx 0.0278 \\ 3 \longrightarrow 2/36 \approx 0.0556 \\ 4 \longrightarrow 3/36 \approx 0.0833 \\ 5 \longrightarrow 4/36 \approx 0.1111 \\ 6 \longrightarrow 5/36 \approx 0.1389 \\ 7 \longrightarrow 6/36 \approx 0.1667 \\ 8 \longrightarrow 5/36 \approx 0.1389 \\ 9 \longrightarrow 4/36 \approx 0.1111 \\ 10 \longrightarrow 3/36 \approx 0.0833 \\ 11 \longrightarrow 2/36 \approx 0.0556 \\ 12 \longrightarrow 1/36 \approx 0.0278 \end{cases} \qquad (2.17)$$

Note that this probability distribution is *normalized*:

$$\sum_{n=2}^{12} w_S(n) \ = \ 1 \ . \qquad (2.18)$$

2.5 Expectation value

Not all values which a random variable can take are equally probable. This is in fact what the probability distribution tells us! It thus makes sense to ask, which is the "most likely" value which the random variable will take. This leads to the notion of the **expectation value**.

The expectation value is defined as the sum over all possible outcomes of the random variable, **weighted with the probability of their occurrence**. Hence we get the following formula:

$$\langle X \rangle \ = \ \sum_x x w_X(x) \ . \qquad (2.19)$$

The angular brackets $\langle \cdots \rangle$ always denote the average ("weighting") over the probability distribution.

In the case of the random variable S defined above, we would calculate the expectation value like this:

$$\begin{aligned} \langle S \rangle \ &= \ \sum_{n=2}^{12} n \, w_S(n) \\ &= \ 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\ &= \ 7 \ . \end{aligned}$$

Hence: If you throw with two dice, you will (on average) expect that the sum is 7. In this particular case 7 is also the most likely value of the probability distribution (i.e., where it has its maximum), **but this need not generally be the case**.

Second example: A friend suggests the following "game": He will flip a coin and we count, how often we get successive "heads" (starting with the first throw!). If we get n heads, he will pay us n Euro. He demands that we pay a price of 2 Euro to him in order to play the game. We could get tremendously rich that way, for just 1 Euro! Should we do this?

Let's first say that N is the random variable which counts how often we had "head" in a row, starting from the first throw. What is the probability distribution of N? Evidently, the probability for getting no head at all ($N = 0$, the series right away starts with a tail) is $1/2$. So with probability $1/2$ we don't get anything. We will get exactly 1 head if the first throw is a head and the second is a tail, so the probability is $1/4$. Hence, the probability for exactly n heads in a row at the beginning is

$$w_N(n) \; = \; \frac{1}{2^{n+1}} \; . \tag{2.20}$$

This is normalized:

$$\sum_{n=0}^{\infty} w_N(n) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \; = \; 1 \; . \tag{2.21}$$

And the expectation value therefore is

$$\begin{aligned} \langle N \rangle \; &= \; \sum_{n=0}^{\infty} n \, w_N(n) \\ &= \; \sum_{n=0}^{\infty} n \, 2^{-(n+1)} \\ &= \; 0\frac{1}{2} + 1\frac{1}{4} + 2\frac{1}{8} + 3\frac{1}{16} + 4\frac{1}{32} + \cdots \\ &\vdots \\ &= \; 1 \end{aligned} \tag{2.22}$$

We thus see, on average, we can only expect to get 1 Euro in a game of this kind. But we have to pay 2 Euro in order to play it! So, on average, we'll just lose 1 Euro. We therefore should not play.

Note that in this particular case the expectation value of the random variable X is 1, but its most likely value is 0 (i.e., there it has its maximum). Hence, the position of the maximum of $w_X(x)$ and the expectation value need not be the same. Well, otherwise the definition of the expectation value would not be terribly interesting!

2.6 Variance

The expectation value tells us which average value to expect, but it does not tell us how far off in each particular instance of a random experiment we would be. In order to have a knowledge of the "spread" of the distribution, we will define the **variance**.

But first of all, let us define the n^{th} moment of a random variable X as follows:

$$\langle X^n \rangle \; = \; \sum_x x^n w_X(x) \tag{2.23}$$

The variance σ_X^2 is then defined as the difference between the second moment and the square of the first moment:

$$\sigma_X^2 \ := \ \langle X^2 \rangle - \langle X \rangle^2 \, . \tag{2.24}$$

Note that this can also be written like this:

$$\sigma_X^2 \ := \ \left\langle \left(X - \langle X \rangle \right)^2 \right\rangle \, . \tag{2.25}$$

Proof:

$$\left\langle \left(X - \langle X \rangle \right)^2 \right\rangle \ = \ \left\langle X^2 - 2X\langle X \rangle + \langle X \rangle^2 \right\rangle \ = \ \langle X^2 \rangle - 2\langle X \rangle \langle X \rangle + \langle X \rangle^2 \ = \ \langle X^2 \rangle - \langle X \rangle^2 \, .$$

Equation (2.25) tells us that the variance is the **average of the square of the difference between the random variable and its average**, or short, the **mean square**.

The square root of the variance is called the **standard deviation**. This is the value most often cited as an error in a measurement. (Note that a measurement is also some kind of random variable: The outcome is not always the same, it fluctuates a bit due to tiny random effects we have no control over, and thus it has an expectation value and some spread around it.)

Example: What is the variance and the standard deviation of the random variable S defined above, the sum of the numbers of two dice? Solution: We first need to calculate $\langle S^2 \rangle$:

$$\begin{aligned} \langle S^2 \rangle \ &= \ \sum_{n=2}^{12} n^2 \, w_S(n) \\ &= \ 4\frac{1}{36} + 9\frac{2}{36} + 16\frac{3}{36} + 25\frac{4}{36} + 36\frac{5}{36} + 49\frac{6}{36} + 64\frac{5}{36} + 81\frac{4}{36} + 100\frac{3}{36} + 121\frac{2}{36} + 144\frac{1}{36} \\ &= \ 54\frac{5}{6} \end{aligned} \tag{2.26}$$

Hence, the variance is

$$\sigma_S^2 \ = \ \langle S^2 \rangle - \langle S \rangle^2 \ = \ 54\frac{5}{6} - 7^2 \ = \ 5\frac{5}{6} \, , \tag{2.27}$$

and the standard deviation is

$$\sigma_S \ = \ \sqrt{5\frac{5}{6}} \ \approx \ 2.415 \, . \tag{2.28}$$

-