Mathematical Basics for Physical Chemistry

Fabian M. Faulstich, PhD

Sept. 7, 2020

1 Derivation

We briefly recall that a function f from a set X to a set Y is defined by a set G of ordered pairs (x, y) such that $x \in X, y \in Y$, and every element of X is the first component of exactly one ordered pair in G. In other words, for every element x in X, there is exactly one element y in Y such that the ordered pair (x, y) belongs to the set of pairs defining the function f. The sets X and Y are very general, meaning that they can either be single or vector valued.

Example 1.1 *The following examples are single and vector valued functions:*

(i) $f_1 : [0,1] \to [0,1]\ ;\ x \mapsto x$

(ii) $f_2 : [0,2] \to [0,4]\ ;\ x \mapsto x^2$

(iii) $f_3 : [0,1] \times [0,1] \to [-1,1]\ ;\ \binom{x}{y} \mapsto x^2 - y^2$

(iv) $f_4 : [0,1] \times [0,1] \to [0,1] \times [-1,0]; \binom{x}{y} \mapsto \binom{x}{-y}$

The **derivative** of a (real valued) function measures the *sensitivity* to change of the function value with respect to a change in its argument. It defines the change of a function in a single point. This can be expressed through the following limit process:
Consider the function $f : X \to Y$, then for $x_0 \in X$ we define

$$\begin{aligned} f'(x_0) = \frac{df}{dx}(x_0) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \\ \Leftrightarrow 0 &= \lim_{x \to x_0} \frac{|f(x) - (f(x) + f'(x_0)(x - x_0))|}{|x - x_0|} \end{aligned} \tag{1}$$

One of the most important application of one-dimensional derivation is the description of motion: Consider the position of an object in motion on a line of length L. This can described by a function of time, since the object may change location as times proceeds. We call this function $s : [0, T] \to [0, L], t \mapsto s(t)$. The speed (velocity) v of this object is defined as the rate of change of its position s over time, i.e.,

$$v(t_0) = \frac{ds}{dt}(t_0) = \dot{s}(t_0). \tag{2}$$

Note that the above describes the derivative of function in one variable, namely, $x \in X$. Moreover, the acceleration a of this object is described by the rate of change of its speed v, i.e.,

$$a(t_0) = \frac{dv}{dt}(t_0) = \dot{v}(t_0) = \frac{d^2s}{dt^2}(t_0) = \ddot{s}(t_0). \tag{3}$$

There are three rules in connection with derivation that become handy when deriving analytical results:

(i) The product rule:

$$\frac{d}{dx}\left(f(x)g(x)\right) = \left(\frac{d}{dx}f(x)\right)g(x) + f(x)\left(\frac{d}{dx}g(x)\right) \tag{4}$$

(ii) The chain rule:

$$\frac{d}{dx}\left(f(g(x))\right) = \frac{d}{dy}f(g(x))\frac{d}{dx}g(x) \tag{5}$$

(iii) The quotient rule:

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{\left(\frac{d}{dx}f(x)\right)g(x) - f(x)\left(\frac{d}{dx}g(x)\right)}{g^2(x)} \tag{6}$$

Example 1.2 *The following are different examples, illustrating the application of the product, quotient and chain rule of differentiation:*

(i) Consider $f(x) = 3x^2 + x$ and $g(x) = x^5$, then

$$\begin{aligned}
\frac{d}{dx}f(x)g(x) &= (6x+1)x^5 + 5x^4(3x^2+x) = 21x^6 + 6x^5 \\
\frac{d}{dx}\frac{f(x)}{g(x)} &= \frac{(6x+1)x^5 - 5x^4(3x^2+x)}{x^{10}} = -(9x^{-4} + 4x^{-5}) \\
\frac{d}{dx}\left(f(g(x))\right) &= (6x^5+1)5x^4 = 30x^9 + 5x^4
\end{aligned} \tag{7}$$

(ii) Consider $f(x) = \exp(-x^2)$ and $g(x) = \sin(1/x)$, then

$$\begin{aligned}
\frac{d}{dx}f(x)g(x) &= -2x\exp(-x^2)\sin(1/x) - \exp(-x^2)\frac{\cos(1/x)}{x^2} \\
&= -\frac{\exp(-x^2)}{x^2}(2x^3\sin(1/x) + \cos(1/x)) \\
\frac{d}{dx}\frac{f(x)}{g(x)} &= \frac{-2x\exp(-x^2)\sin(1/x) + \exp(-x^2)\frac{\cos(1/x)}{x^2}}{\sin^2(1/x)} \\
&= -\frac{\exp(-x^2)}{\sin(1/x)x^2}(2x^3\sin(1/x) + \cos(1/x)) \\
\frac{d}{dx}\left(f(g(x))\right) &= \frac{2\sin(1/x)\cos(1/x)}{x^2}\exp(-\sin^2(1/x))
\end{aligned} \tag{8}$$

In order to extend the definition of the derivative to the more general case of higher-dimensional derivatives, we observe the following: The derivative itself is a function, namely, for each, $x_0 \in X$ the function f' yields one single value, i.e., a constant. Note that this does not mean that the function f' is constant itself, just consider $f : [0,1] \to [0,1]\ ;\ x \mapsto x^3$. Clearly, $f' : [0,1] \to [0,1]\ ;\ x \mapsto 3x^2$ is not linear, yet, for any $x_0 \in [0,1]$ the map f' merely yields a single value. Although this observation may seem trivial in one dimension, since it arises directly from the definition of a function, it is key when thinking about vector valued functions: One way to think of the generalization of a *constant* in higher dimensional spaces are matrices. Recall that a matrix is a rectangular array of numbers that can be real or complex valued.

Example 1.3 *The following are a few examples of matrices:*

(i) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (ii) $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ (iii) $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ (iv) $\begin{pmatrix} 5 \end{pmatrix} = 5$ (v) $\begin{pmatrix} 1 & 4 \\ 2 & -\frac{1}{2} \\ \frac{3}{8} & 7 \end{pmatrix}$

Recall also that every matrix defines a linear function. Now, we put this knowledge together and take a look at higher dimensional derivatives: We consider the function $f : X \to Y$ where X and Y are vector spaces equipped with the norms (i.e. measures of distances) $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. The generalization of the differential quotient then reads:

$$0 = \lim_{x \to x_0} \frac{\|f(x) - (f(x_0) + Df(x_0)\,(x - x_0))\|_Y}{\|x - x_0\|_X} \tag{9}$$

We call the linear function (i.e. the matrix) $Df(x_0) : X \to Y$ the derivative of f in x_0. Note that the derivative of f is not necessarily linear(!), moreover, it is a function that maps an element $x_0 \in X$ onto a matrix, i.e., $Df : X \to L(X,Y)$, where $L(X,Y)$ is simply the space of linear functions (matrices) that map from X to Y. The derivative Df is also called the Jacobian, and can be expressed by the directional derivatives. We define

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t} \tag{10}$$

as the partial derivative of f in direction e_i, where e_i is the i-th basis vector of X. Using this notation we can write out the Jacobian as

$$Df = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{11}$$

Example 1.4 *We consider the following functions and higher dimensional derivatives*

(i) Consider $f : \mathbb{R}^3 \to \mathbb{R}^3$; $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ -2x_2 \\ x_1 + 2x_2x_1 - x_3^3 \end{pmatrix}$ *the Jacobian is given by:*

$$Df(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 1 & 2x_2 & -3x_3^2 \end{pmatrix}$$

(ii) Consider $f : \mathbb{R}^2 \to \mathbb{R}^3$; $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1^2 + x_2^2 \\ x_1x_2 \\ x_1^3 + x_2^3 \end{pmatrix}$ *the Jacobian is given by:*

$$Df(x) = \begin{pmatrix} 2x_1 & 2x_2 \\ x_1 & x_2 \\ 3x_1^2 & 3x_2^2 \end{pmatrix}$$

(ii) Consider $f : \mathbb{R}^3 \to \mathbb{R}$; $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \left(x_1^3 + x_2^2x_3 - x_3\right)$ *the Jacobian is given by:*

$$Df(x) = \begin{pmatrix} 3x_1^2 & 2x_2x_3 & x_2^2 - 1 \end{pmatrix} = \left(\nabla f(x)\right)^T$$

where we introduced the nabla operator ∇ *describing the gradient of a single valued function.*

An important application of the derivative is the numerical solution of nonlinear equation, namely, through the Newton–Raphson method often simply referred to as Newton's method. Consider systems of k (nonlinear) equations, i.e., $f(x_*) = b$ where $x_*, b \in \mathbb{R}^k$. This amounts to finding the zeroes of the function $g : \mathbb{R}^k \to \mathbb{R}^k; x \mapsto f(x) - b$. Recall, the idea of the Newton iteration (i.e, applying the Newton–Raphson method) is to start with an initial guess which is reasonably close to the true root, then approximate the function by its tangent line and compute the x-intercept of this tangent line. More formally, suppose g is a differentiable function, and we have some current approximation x_n. The equation of the tangent line to g at $x = x_n$ is

$$y = Dg(x_n)\,(x - x_n) + g(x_n). \tag{12}$$

The x-intercept of this line (the value of x which makes $y = 0$) is taken as the next approximation, x_{n+1}, to the root, so that the equation of the tangent line is satisfied when

$$0 = Dg(x_n)\,(x - x_n) + g(x_n). \tag{13}$$

Solving for x_{n+1} yields

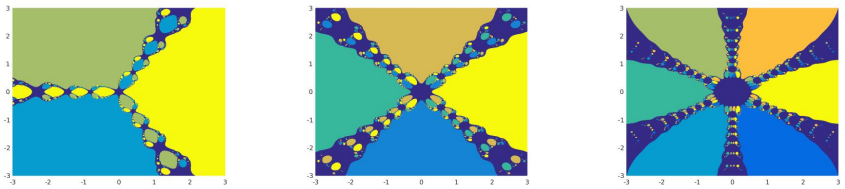
$$x_{n+1} = x_n - Dg(x_n)^{-1}g(x_n). \tag{14}$$

Rather than actually computing the inverse of the Jacobian Dg, which can be numerically **very** costly, one can equivalently solve the system of linear equations

$$Dg(x_n)\tilde{x}_n = -g(x_n) \tag{15}$$

and obtain x_{n+1} through $x_{n+1} = \tilde{x}_n + x_n$.

Recall that the Newton–Raphson method is very sensitive with respect to the starting point. In fact, slight deviation of your starting point can have severe impact on what solution is found. This can be visualized by means of the newton fractals that show the complicated dependence of the method's solution and the starting point (see Fig. 1).



(a) (b) (c)

Figure 1: Different Newton fractals:
(a) $g(z) = z^3 - 1$
(b) $g(z) = z^4 - 1$
(c) $g(z) = z^6 - 1$

2 Integration

Integration is concerned with the area under a single valued function. An example for when this quantity is of interest is the expression of work in a closed system, i.e.,

$$w = -\int_{V_i}^{V_f} p_{\text{ex}}(V)dV \ . \tag{16}$$

Note that the area, that we are concerned with, can be restricted to a region that is a simple interval $[a, b]$ or the union of different intervals $[a_1, b_1] \cup [a_2, b_2]$ or even the entire space, e.g., $\mathbb{R}$. We are here concerned with the so called Riemann integral that is based on a partitioning of the domain of f or merely a subregion of it that we are interested in. For sake of simplicity we consider the interval $[a, b]$ to be of interest for us. A possible partitioning can be obtained a follows: Given a value δ, we define $\{x_0^{(\delta)}, x_1^{(\delta)}, ..., x_n^{(\delta)}\}$, where $x_0^{(\delta)} = a$, $x_n^{(\delta)} < b$ and $x_i^{(\delta)} - x_{i-1}^{(\delta)} = \delta$. Said partition has $N_\delta = \lfloor (b-a)/\delta \rfloor \in \mathbb{N}$ elements, where $\lfloor \cdot \rfloor$ denotes rounding down to the closest integer. The Riemann sum that corresponds to this particular partitioning is given by

$$\mathcal{R}f[\delta] = \sum_{i=0}^{N_\delta} f(x_i^{(\delta)})\delta. \tag{17}$$

The above described partition can be defined for any value of δ. This allows us to define the following sequence of Riemann sums: Consider $\delta_n = 1/n$ and $N_n = \lfloor (b-a)n \rfloor$, then

$$\mathcal{R}_n = \mathcal{R}f[1/n] = \sum_{i=0}^{N_n} \frac{f(x_i^{(n)})}{n} \tag{18}$$

defines a sequence $(\mathcal{R}_n)_{n\in\mathbb{N}}$ where the partition $\{x_i^{(n)}\}_{i=0}^{N_n}$ is constructed as described above. An illustration of the above procedure is given by Figure 2.

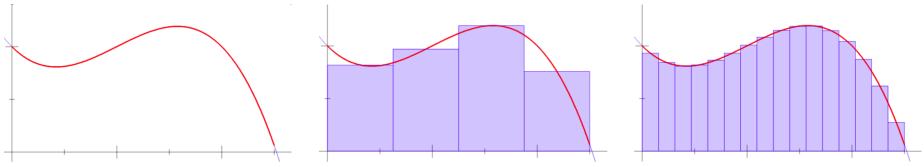


Figure 2: From left to right: the considered function with region of interest, partition of the region for $n = 4$, partition of the region for $n = 16$. Note that the illustration differs slightly from the described procedure since the function is here evaluated at the midpoint of the interval $[x_i^{(n)}, x_{i+1}^{(n)}]$ for $i \in \{0, ..., n\}$ and where $x_{n+1}^{(n)} = b$.

The limit of the sequence $(\mathcal{R}_n)_{n\in\mathbb{N}}$ (if it exists!) is then called the Rieman

integral, or simple the integral of f over the interval $[a, b]$, and we write

$$\int_a^b f(x)dx = \int_{[a,b]} f(x)dx = \lim_{n\to\infty} \mathcal{R}_n = \lim_{n\to\infty} \sum_{i=0}^{N_n} \frac{f(x_i)}{n}. \tag{19}$$

The consideration of Riemann sums is of very high practical relevance, since it yields a first, but yet very direct approach to the numerical approximation, i.e., on a computer, to evaluate complicated integral expressions.

Another direct application of Riemann sums, is given by numerical methods for solving differential equations. Assume the following Cauchy problem (differential equation with a given boundary condition):

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0 = 0) = 1 \end{cases} \tag{20}$$

We know the solution to this problem is $y(t) = e^t$, yet we also know that this differential equation can be transformed to the following integral form:

$$y(t) = \int_{t_0=0}^{t} y'(s)ds + y(t_0 = 0) = \int_{t_0=0}^{t} f(s, y(s))ds + y(t_0 = 0) \tag{21}$$

Assuming that we are interested in $y(t_0+h)$, with h sufficiently small. Then, we can use the trivial Riemann sum, i.e., approximating the integral by a square, and find

$$y(t_0 + h) \approx f(t_0, y(t_0))h + y(t_0). \tag{22}$$

We can furthermore continue adding small increments of h to $t_n = t_0 + nh$, and find the recursive form

$$y(t_{n+1}) = y_n + hf(t_n, y(t_n)), \tag{23}$$

which is known as the explicit Euler method for solving differential equations. Similarly, we can derive the implicit Euler, the Crank–Nicolson as well as other numerical methods for solving differential equations by taking different approximations of the integral $\int_{t_0=0}^{t} f(s, y(s))ds$.

For analytical, i.e., 'on paper', results the fundamental theorem of calculus is the most relevant results. This theorem consists of two basic parts:

(i) Let f be a continuous real valued function defined on a closed interval $[a, b]$. Let F be the function defined, for all x in $[a, b]$, by

$$F(x) = \int_a^x f(t)dt. \tag{24}$$

Then F is uniformly continuous on $[a, b]$ and differentiable on the open interval (a, b), and

$$F'(x) = f(x) \tag{25}$$

for all x in (a, b). The function F is also referred to as an antiderivative.

(ii) Let f be a real-valued function on a closed interval $[a, b]$ and F an antiderivative of f in $[a, b]$. If f is Riemann integrable on $[a, b]$ then

$$\int_a^b f(x)dx = F(b) - F(a). \tag{26}$$

Example 2.1 *The following are different examples that illustrate the fundamental theorem of calculus introducing several antiderivatives:*

(i) Consider $f(x) = x^3 + x/2$, then

$$\begin{aligned} \int_0^1 f(x)dx &= \Big[\frac{1}{4}(x^4 + x^2)\Big]_0^1 = \frac{1}{4}(2 - 0) = \frac{1}{2} \\ \int_2^4 f(x)dx &= \Big[\frac{1}{4}(x^4 + x^2)\Big]_2^4 = \frac{1}{4}(256 + 16 - 16 - 4) = 63 \end{aligned} \tag{27}$$

(ii) Consider $f(x) = 3x^2 \exp(x^3)$, then

$$\int_0^1 f(x)dx = \Big[\exp(x^3)\Big]_0^1 = e - 1 \tag{28}$$

(iii) Consider $f(x) = 1/x$, then

$$\int_e^{\exp(e)} f(x)dx = \Big[\ln(x)\Big]_e^{\exp(e)} = e - 1 \tag{29}$$

Moreover, there are two other, very important results in connection with integration:

(i) Integration by parts:

$$\begin{aligned} \int_a^b u(x)v'(x)dx &= \Big[u(x)v(x)\Big]_a^b - \int_a^b u'(x)v(x)dx \\ &= u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)dx. \end{aligned} \tag{30}$$

More compactly,

$$\int u\,dv \;=\; uv - \int v du. \tag{31}$$

(ii) Integration by substitution:
Let $\varphi : [a, b] \to I$ be a differentiable function with a continuous derivative, where $I \subseteq \mathbb{R}$ is an interval. Suppose that $f : I \to \mathbb{R}$ is a continuous function. Then, setting $u = \varphi(x)$

$$\int_a^b f(\varphi(x))\varphi'(x)\,dx = \int_{\varphi(a)}^{\varphi(b)} f(u)\,du. \tag{32}$$

Example 2.2 *The following are different examples, illustrating integration by parts and by substitution:*

(i) We will use integration by parts:

$$\int x\cos(x)dx = x\sin(x) - \int cos(x)dx = x\sin + \sin(x) + c \tag{33}$$

(ii) We will use integration by parts:

$$\begin{aligned} \int e^x \sin(x)dx &= e^x \sin(x) - \int e^x \cos(x)dx \\ &= e^x \sin(x) - e^x \cos(x) - \int e^x \sin(x)dx \\ \Leftrightarrow \int e^x \sin(x)dx &= \frac{\sin(x) - \cos(x)}{2} e^x + c \end{aligned} \tag{34}$$

(iii) We will use integration by substitution:

$$\begin{aligned} \int_0^{\sqrt{\pi/2}} \cos(x^2)6xdx = 3\int_0^{\sqrt{\pi/2}} \cos(x^2)2xdx &\underset{\varphi(x)=x^2}{=} 3\int_0^{\pi/2} \cos(u)du \\ &= 3\sin(u)\Big|_{u=0}^{\pi/2} = 3 \end{aligned} \tag{35}$$

(iv) We will use integration by substitution:

$$\begin{aligned} \int_0^{\sqrt{e-1}} \frac{x}{x^2+1}dx = \frac{1}{2}\int_0^{\sqrt{e-1}} \frac{2x}{x^2+1}dx &\underset{\varphi(x)=x^2+1}{=} \frac{1}{2}\int_1^e \frac{1}{u}du \\ &= \frac{1}{2}\ln(u)\Big|_{u=1}^{e} = \frac{1}{2} \end{aligned} \tag{36}$$

3 Probability Distributions

In this section we want to get a better understanding of the framework that allows us to describe probabilistic experiment. One specific application can be found in statistical mechanics, where the velocity of individual particles cannot be calculated deterministically. Probabilistically, this can be modeled by means of the Maxwell–Boltzmann distribution. The goal of this section is to get a better understanding of the notions of the underlying probability theory. To that end, we need to get a rough idea of the basic concepts of probability theory, which is an entire field in itself. Subsequently, we will briefly introduce the ideas of:

1) Probability triples
2) Random variables
3) Probability densities
4) Stochastic moments
   - Expectation values
   - Variance

In order to fully understand the above concepts we need to full power of measure theory: an event in probability is a measurable set, a random variable is a measurable function on the sample space, the expectation of a random variable is its integral with respect to the probability measure, and so on. However, we will try to maneuver through the basics without going into too much detail, but yet enough intuition in order to understand what kind of probabilistic objects are considered in the course on physical chemistry.

1) Probability triple: A model for an experiment involving randomness takes the form of a probability triple, which consists of three components:

- the sample space Ω
- the event space $\mathcal{F}$
- the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$

We will try to illustrate the different objects and their relevance using some examples.

Sample space: The first thing we would like to do is to characterize the set of possible outcomes of our experiment. This is called the **sample space**, and is commonly denoted by Ω.
Examples:

(i) Consider the toss of a fair die, then we can observe the appearance of an integer between one and six, i.e., $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(ii) Consider the toss of coin, then outcome can be ether heads or tails, i.e., $\Omega = \{0, 1\}$, where one means heads and zero means tails.

(iii) Drawing a card of a deck (52 cards), we can assign every single card a unique integer between one and 52, i.e., $\Omega = \{1, ..., 52\}$.

(iv) Measuring the velocity of a single particle in a gas, in principle, any positive real number can be possible, i.e., $\Omega = \mathbb{R}_+$

Event space: Next we are characterizing the collection of outcome, i.e., all set of possible outcomes. This is called the **even space**, and is commonly denoted by $\mathcal{F}$. The event space is different from the sample space as it describes the collections of outcomes. An example for an event would be that we get either 1,2 or 4 as outcome of the toss of the die. This already hints that in the discrete case, it is always possible to choose the power set of the sample space Ω as event space, i.e., $\mathcal{F} = \mathcal{P}(\Omega)$. We recall that the power set of any set is the set of all subsets. In the continuous case the situation is a bit more complected. We need to ensure that our events are measurable, meaning, that we can assign a value through a function that has very specific properties, namely, a measure. A detailed description of measurable sets goes far beyond the scope of this introduction. In order to conclude the discussion on event spaces we write them out for the considered examples:

(i) Toss of a coin, i.e., $\Omega = \{1, 2\}$. Then, $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

(iv) Measuring the velocity of a single particle in a gas, i.e., $\Omega = \mathbb{R}_+$. Then, $\mathcal{F} = \mathfrak{L}(\mathbb{R}_+)$, i.e., the Lebesgue-measurable sets of $\mathbb{R}_+$.

Probability measure: Next, we want to assign probabilities of events. To that end, we require a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, i.e., a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ fulfilling:

i) $\mathbb{P}(\omega) \geq 0$ for all $E \in \mathcal{F}$ (Nonnegativity)

ii) $\mathbb{P}(\emptyset) = 0$ (Null empty set)

iii) For all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{F}$

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(E_k) \qquad (\sigma - \text{additivity})$$

iv) $\mathbb{P}(\Omega) = 1$ (Normalization)

In the discrete case, it is always possible to assign such a measure, namely, the counting measure:

$$\mathbb{P}_c : \mathcal{F} \to [0, 1]; \omega \mapsto \frac{|\omega|}{|\Omega|}. \tag{37}$$

In the continuous case, we rely again on results from measure theory, and for now shall assume some one has given us the magical probability measure. In order to conclude the discussion on probability measures we write them out for the considered examples:

(i) Consider the toss of a coin, where $(\Omega = \{1, 2\}, \mathcal{F} = \mathcal{P}(\Omega))$. Then,

$$\mathbb{P} : \mathcal{F} \to [0, 1], \omega \mapsto \mathbb{P}(\omega) = \int_{\omega} d\mathbb{P} = \frac{|\omega|}{|\Omega|} \tag{38}$$

(ii) Measuring the velocity of a single particle in a gas, i.e., $(\Omega = \mathbb{R}_+, \mathfrak{L}(\mathbb{R}_+)$. Then,

$$\mathbb{P} : \mathfrak{L}(\mathbb{R}_+) \to [0, 1], \omega \mapsto \mathbb{P}(\omega) = \int_{\omega} d\mathbb{P} = \int_{\omega} 4\pi \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2kT}} dv \tag{39}$$

2) Random variable: The random variable is arguably the most important concept in probability theory. It describes the considered experiment and carries the information of the outcome's distribution. Formally, we define random variables as functions $X : \Omega \to \mathbb{R}$. The probability that a certain event S will occur, is give by the following expression:

$$\mathbb{P}(X \in S) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in S\}) = \int_{X^{-1}(S)} d\mathbb{P} \tag{40}$$

3) Probability density: A probability density is connected to a certain random variable following a given distribution. Probability densities make our life easier since they allow us to use regular integration, instead of the integration with respect to a given probability measure. The key is again given by measure theory, which provides us with the proper tool to retract the integral to the Lebesgue measure (which in our cases coinsides with the Riemann integration), namely, the Radon–Nikodym derivative. We obtain this though the following formal manipulation:

$$\mathbb{P}(X \in S) = \int_{X^{-1}(S)} d\mathbb{P} = \int_S dX_* \mathbb{P} = \int_S \frac{dX_* \mathbb{P}}{d\lambda} d\lambda = \int_S f_X d\lambda \tag{41}$$

4) Stochastic moments: The n-th moment of a random variable X following a given distribution with corresponding density f_X is given by

$$\mu_n = \int_{-\infty}^{\infty} x^n f_X(x)\, dx. \tag{42}$$

Second and higher order moments can be centralized, meaning they take the shape

$$\mu_{n,c} = \int_{-\infty}^{\infty} (x - \mu_0)^n f_X(x)\, dx. \tag{43}$$

The most important stochastic moments are the first moment and the second centralized moment which correspond to the expectation value and the variance, respectively. Note that in the case of the variance (from now on denoted $\mathbb{V}$), the following holds

$$\begin{aligned}\mathbb{V}(X) &= \int (x - \mathbb{E}(X))^2 f_X(x)\, dx = \int (x^2 - 2\mathbb{E}(X)x + \mathbb{E}(X)^2 f_X(x)\, dx \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2\end{aligned} \tag{44}$$

Example:

(i) Given the random variable X that follows the Bernoulli distribution, i.e., it has two possible outcome $\Omega = \{1, 0\}$, with probability p and $q = 1 - p$, respectively. The probability density is given by

$$f_X(x) = p\delta(x - 1) + q\delta(x - 0), \tag{45}$$

where δ is given by

$$\int_{\mathbb{R}} g(x)\delta(x - x_0)dx = g(x_0). \tag{46}$$

Then, the first and second stochastic moments are given by

$$\begin{aligned}\mathbb{E}(X) &= \int_{\mathbb{R}} x f_X(x)dx = p \cdot 1 + q \cdot 0 = p \\ \mathbb{E}(X^2) &= \int_{\mathbb{R}} x^2 f_X(x)dx = p \\ \Rightarrow \quad \mathbb{V}(X) &= p - p^2 = p(1 - p) = pq\end{aligned} \tag{47}$$

(ii) Given the random variable X that follows the Maxwell–Boltzman distribution, i.e., the probability density is given by

$$f_X(x) = av^2 e^{-bv^2} \mathbb{1}_{[0,\infty)}(x), \tag{48}$$

where $\mathbb{1}_{[0,\infty)}(x)$ is the indicator function on $[0, \infty)$ returning one if $x \in [0, \infty)$ and zero else. Then, the first stochastic moment is given by

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty av^3 e^{-bv^2} dv = \frac{1}{2}\int_0^\infty av^2 e^{-bv^2} 2v dv \\ &\underset{\varphi(v)=v^2}{=} \frac{1}{2}\int_0^\infty aue^{-bu} du \\ &= -\frac{a}{2b} ue^{-bu}\Big|_{u=0}^\infty + \frac{a}{2b}\int_0^\infty e^{-u} du = -\frac{a}{2b}\end{aligned} \tag{49}$$