Radboud University Nijmegen

The Method of Arbitrary Functions and Its Philosophical Relevance

Vera Bosveld

Bachelor Thesis in Mathematics

Supervisor: prof. dr. Klaas Landsman

Second reader: prof. dr. Henk de Regt

Contents

1 Introduction

How does one derive probability? Or, differently put, how does one *interpret* probability? Within the debate on probability there exist different interpretations of probability, and each of them has its own perspective on how to perceive it. Within this debate, one is not only interested in the mathematical foundation on which probability should be build; it concerns the metaphysics of probability and studies its reality. Is probability something that really exists out there or is it something humans have constructed to describe situations we *ourselves* are uncertain about; is it objective or subjective? Many philosophers have racked their brains over this and it still remains a central question in the debate on probability. Putting this question aside for now, let us review some interpretations of probability a bachelor student of mathematics has encountered and find out why these are not universally accepted as *the* interpretation of probability.

The first one goes by the name *finite frequentism*, mostly seen in statistics. A finite frequentist equates probability to the relative frequency with which a certain outcome occurs within a finite number of trials. Suppose that someone flips a fair coin 1000 times and 500 of those result in heads, a frequentist would then assign probability $\frac{1}{2}$ to the event that the coin lands heads. Evidently, most experiments of 1000 coin flips do not result in heads exactly 500 times. What if, by some accident, heads faces the sky a mere 100 times. Identifying probability with relative frequencies, one is obligated to assign a probability of $\frac{1}{10}$ to the event. Moreover, if such an experiment were to be executed multiple times, then it would probably give all kinds of relative frequencies. Is the probability of the event then equal to all these different values? And what if the coin is tossed only once? Consequently, the probability of heads can only be equal to 0 or 1. Similarly, each time a fair coin is tossed an odd number of times, the possibility that the probability of heads is equal to one half does not even exist. I can imagine that at this point one might say that the connection between probability and frequentism is not *this* close and should be loosened up. Perhaps that relative frequency should pose as evidence for probability, not as probability itself. But that brings us right back to where we started; if frequencies do not define probability, then what does?

What about the classical interpretation of probability? Define the sample space Ω which consists of all possible outcomes to a certain experiment. Moreover, *assume* that all of these outcomes $\omega \in \Omega$ are equally likely to happen. The probability of a certain event E is then defined as:

$$P(E) = \frac{\#\{\omega \in \Omega : \omega \text{ leads to } E\}}{\#\Omega}.$$

For the coin toss, setting $\Omega = \{heads, tails\}$ is one way of defining the sample space. When questioning the probability of heads, it seems that the answer is already presumed by the definition, because all elements in Ω, heads and tails, are assumed to be equally likely to happen. Thus, we see that this interpretation requires prior knowledge on how to interpret probability. Although this interpretation proves useful in games of chance, such a circular definition does not hold up well in metaphysics.

The fact that these two widely used concepts of probability face some issues should hopefully give an idea of the difficulty of interpreting probability. What if, instead of using frequencies or sample spaces, one uses the dynamical properties of a chance experiment to interpret probabilities? This is what happens in the *method of arbitrary functions*. Within this method, one finds all the factors that determine the end result. So if, for example, someone flips a coin, then the exact way that this coin leaves this person's hand (spin speed, upward speed etc.) determines the outcome; heads or tails. These factors are called the *initial conditions* or *initial values*. So, if we know the exact initial conditions, then we would know whether the coin lands heads or tails. However, we do not know these exact values. For example, when I flip a coin, I cannot know for sure with how much force I am going to flip the coin, let alone knowing with how much force someone else flips a coin.

We see that the uncertainty over the outcome results from the uncertainty over the initial conditions. Or, mathematically expressed, the probability p of a certain event A depends on the distribution of the initial conditions, often called the *initial distribution*. However, what stands out about the method of arbitrary functions is that it limits this dependency quite remarkably. That is, it provides a probability that is *almost* independent of the initial distribution. Consequently, it is *almost* solely dependent on the dynamical properties of the experiment; on something objective. The near objectivity of this probability provides fruitful ground for an alternative interpretation of probability. Therefore, many philosophers have attempted to construct a *new* interpretation of probability inspired by the method of arbitrary functions.

In this thesis, I will demonstrate both the mathematical and philosophical side of the method of arbitrary functions. In Chapter 2, I will give an introduction to the method using the coin toss as the main example. Theorem 2.2 demonstrates the method of arbitrary functions particularly applied to the coin toss. In Chapter 3 more general results follow, culminating in the main result, Corollary 3.15. This corollary is the main result because of its capacity to apply the method of arbitrary functions to a variety of deterministic mechanisms, whereas Theorem 2.2 is only applicable to the coin toss (of a certain kind). The last chapter, Chapter 4, concerns the philosophical relevance of the method. As mentioned above, the method of arbitrary functions inspires a new interpretation of probability. In particular, this thesis focuses on the interpretations that Rosenthal and Strevens have constructed. This interpretation contains both promising features and controversial elements, and discussing all of these would be beyond the scope of this thesis. To avoid getting lost in the spider web that is the debate on probability, I will only briefly touch on its promising features, and mainly focus on one particular controversy. That is, the probability interpreted in this way is not entirely independent of the initial distribution. How do Rosenthal and Strevens deal with the existence of certain eccentric initial distributions that yield deviant probabilities and which undermine this interpretation? That is the question that I will answer in Chapter 4.

2 The Coin Toss

We are all familiar with the coin toss and one might question the need to devote a whole chapter to it. The idea is to unfold the method of arbitrary functions using the coin toss as our main example. In doing so, we find that the coin toss might not be as trivial as one would expect. In the first section of this chapter, we will take a look at why the coin toss is a probabilistic affair in the first place. In Section 2.2, we will determine which factors determine the outcome of the coin toss and use these to calculate when exactly the coin lands heads and when it lands tails. In Section 2.3, we will use these calculations to loosely approximate the probability of heads. In doing so, the idea behind the method of arbitrary functions becomes visible. Then, in Section 2.4, more exact results follow, including the main theorem of this chapter (Theorem 2.2). Section 2.5 consists of some notes on the method. And finally, in Section 2.6, we look at other games of chance to which the method can be applied.

2.1 The coin toss as a probabilistic affair

The question whether heads or tails is the outcome of a coin toss actually has an answer when one has the sufficient amount of knowledge. A physicist is capable of predicting the outcome with certainty provided she knows the exact height from which the coin is tossed, the exact 'force' with which the coin is tossed, the exact way the coin rotates, etc. In other words, the outcome of a coin toss is determined in principle. However, people are not capable of knowing exact values and are even less capable of throwing a coin with a certain exactness. So when tossing a coin, the uncertainty over its outcome is simply due to the uncertainty over the way that coin is going to be tossed. We find that the probability of a coin landing heads can be reduced to the probability of a person throwing the coin in such a way that it leads to heads. The "ways" with which a person tosses a coin are called the *initial conditions* or *initial values*. These are the conditions of the beginning of the experiment which completely determine the final result. This reduction to looking at the probability of the initial conditions is the common denominator of all applications of the method of arbitrary functions. This method is therefore only applicable to experiments that are, in principle, deterministic, such as the coin toss, the rolling of a die, the spinning of a wheel and others.

2.2 Physical properties

As mentioned above, the probability of heads reduces to the probability of the initial conditions. Because, if we know what initial conditions lead to heads as outcome *and* we know what the probability is that those initial conditions arise, then we know the probability of heads. So, for this section, we set out to find the set of initial conditions that lead to heads as outcome. We will call this set the *pre-image of heads*.

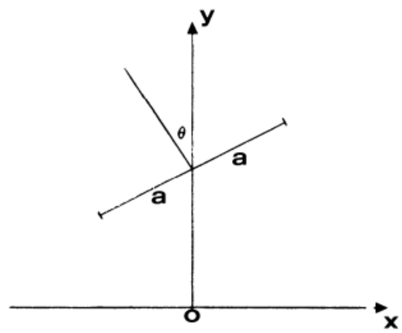

Figure 1: Keller's model of the coin toss ((9), p.192).

In order to do so, we need to discuss which of the million different versions of the coin toss we take as our basis. Having done so, we establish the initial conditions for this particular coin toss. After which we will perform the physics and find the pre-image of heads.

Almost everything we discuss in this section and the next can also be found in Keller's article (9). In his article, Keller uses a simplified representation of the coin toss such that complexity is reduced, which is shown in Figure 1. We reduce complexity by assuming that the coin is a perfect disk with negligible thickness and diameter $2a$. Its center of gravity is at its geometrical center, no air resistance is assumed, heads always faces up right before the coin is tossed, and the center of gravity moves in one dimension; along the y axis. In order to simplify things even more, the coin is assumed to be caught by the palm of someone's hand ($y = 0$), the meaning of this is to eliminate any bouncing the coin might otherwise engage in when touching the surface. Finally, we assume that the coin rotates in two dimensions, so the only rotation parameter we are concerned with is θ. Simplifying the coin toss in this way leaves us with three initial conditions that determine the final outcome; the initial vertical velocity u, the initial angular velocity ω, and the initial height b. Again for simplicity's sake, it is assumed that $b = a$ for each experiment.

Denote the height of the geometrical center of the coin at time t by $y(t)$. Then Newton's equation for the vertical motion of the center of gravity of the coin is

$$\frac{d^2y(t)}{dt^2} = -g. \tag{1}$$

We know $y(0)$ is equal to the initial height a and $y'(0)$ to the initial vertical velocity u. We can therefore solve the differential equation and obtain the following:

$$y(t) = ut - \frac{gt^2}{2} + a. \tag{2}$$

Next the coin is assumed to rotate in two dimensions. This means that it rotates

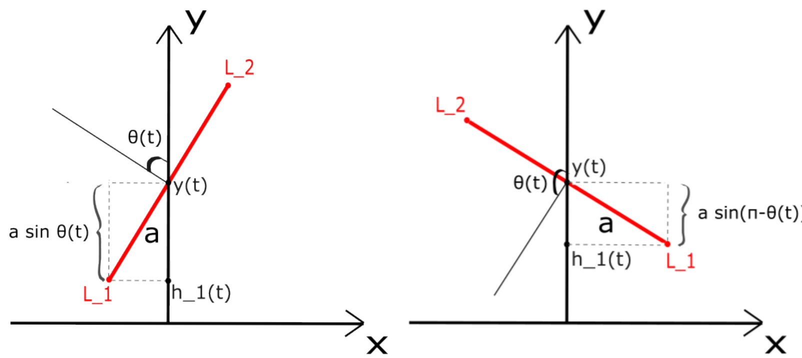

Figure 2: The position of the coin (illustrated as the red line) when $0 \leq \theta(t) \leq \frac{\pi}{2}$ and $\frac{\pi}{2} \leq \theta(t) \leq \pi$, respectively.

around a horizontal axis which lies along the diameter and is parallel to the z axis. Denote the angular position of the coin at time t by the angle $\theta(t)$ between the y axis and the normal to the side of the coin marked heads. Due to the assumption that heads faces up at the beginning of each experiment, we find $\theta(0) = 0$. Additionally, the given initial angular velocity ω gives us $\theta'(0) = \omega$. Furthermore, by our assumption of no air resistance, θ satisfies the following differential equation:

$$\frac{d^2\theta(t)}{dt^2} = 0. \tag{3}$$

All these facts combined gives us a solution for $\theta(t)$:

$$\theta(t) = \omega t. \tag{4}$$

Now suppose the coin touches the surface for the first time at $t = t_0$. Then (2) and (4) are applicable as long $t < t_0$. We have assumed that once $t = t_0$, the coin does not bounce. This means that the side facing up at t_0 is the side facing up when the coin is at rest. In other words, if we calculate which side faces up at t_0, then we know the final outcome. To do so, we write the height of the lowest point of the coin at time t_0 in terms of u and ω, and equate it to 0. Looking at Figure 2, it becomes clear that L_1 is the point of the coin that first touches the surface if $0 \leq \theta(t_0) \leq \pi$ and that the height of L_1 at time t is equal to

$$\begin{aligned} h_1(t) &= \begin{cases} y(t) - a\sin\theta(t), & \text{if } 0 \leq \theta(t) \leq \frac{\pi}{2} \\ y(t) - a\sin(\pi - \theta(t)), & \text{if } \frac{\pi}{2} \leq \theta(t) \leq \pi \end{cases} \\ &= y(t) - a\sin\theta(t). \end{aligned} \tag{5}$$

Similarly, if $\pi \leq \theta(t_0) \leq 2\pi$, then L_2 is the point to first touch the surface. The height of L_2 at time t equals

$$h_2(t) = y(t) + a\sin\theta(t). \tag{6}$$

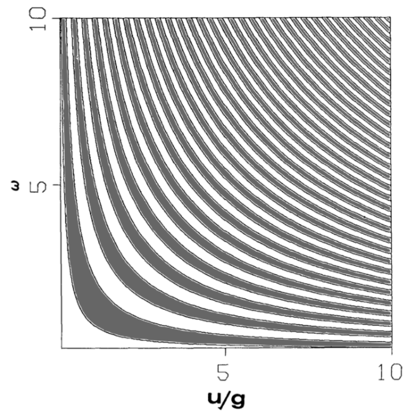

Figure 3: Initial values satisfying (11) are indicated with white, the remaining values are gray ((9), p.193).

And therefore, the height of the lowest point at time t equals

$$h(t) = \begin{cases} y(t) - a\sin\theta(t), & \text{if } 0 \leq \theta(t) \leq \pi \\ y(t) + a\sin\theta(t), & \text{if } \pi \leq \theta(t) \leq 2\pi \end{cases} = y(t) - |a\sin\theta(t)|. \tag{7}$$

Note that $h(t) = 0$ iff $t = t_0$. Furthermore, note that if $(2n - \frac{1}{2})\pi \leq \theta(t_0) \leq (2n + \frac{1}{2})\pi$, then heads is the final outcome of the coin toss. Combining these boundary values with (4) gives us:

$$\omega t_0 = (2n \pm \frac{1}{2})\pi. \tag{8}$$

If $\theta(t_0)$ is equal to either of these boundary values, then $\sin\theta(t_0) = \pm 1$. This, combined with (2) and (7), gives us

$$0 = y(t_0) - a = ut_0 - \frac{gt_0^2}{2} \longrightarrow t_0 = \frac{2u}{g}. \tag{9}$$

Implementing (9) into (8) gives us

$$\omega = (2n \pm \frac{1}{2})\frac{\pi g}{2u}. \tag{10}$$

So we now know that if (u, ω) satisfies

$$\omega \in \frac{\pi g}{2u}[2n - \frac{1}{2}, 2n + \frac{1}{2}], \tag{11}$$

then throwing with that initial velocity u and that angular velocity ω leads to heads as outcome of the coin toss. Figure 3 shows (11) for a few different $n \in \mathbb{N}$.

If $(u/g, \omega)$ falls within a white area of this figure, then those initial conditions lead to heads as outcome. Define:

$$H = \{(u, \omega) : \omega \in \frac{\pi g}{2u}[2n - \frac{1}{2}, 2n + \frac{1}{2}]; \tag{12}$$

$$T = \{(u, \omega) : \omega \in \frac{\pi g}{2u}[2n + \frac{1}{2}, 2(n+1) - \frac{1}{2}]\}. \tag{13}$$

Then H is called the *pre-image of heads* and is equal to the white area in Figure 3. T is called *the pre-image of tails* and is equal to the gray area. As mentioned at the beginning of this section, H is exactly the set we intended to find.

2.3 The probability of heads

Recall that the probability of heads is the probability that those initial conditions arise that lead to heads as outcome. So, by our findings of the previous section we obtain that the probability of heads is:

$$\mathbb{P}_H = \mathbb{P}((u, \omega) \in H) \tag{14}$$

So suppose, for simplicity's sake, that we know that the tosser tosses the coin in a uniformly distributed way over some interval I:

$$I = \{(u, \omega) : u \in [a, a+1], \omega \in [b, b+1]\}, \tag{15}$$

for some $a, b \in \mathbb{R}_{>0}$. Figure 4 shows that if (u, ω) is within the yellow area, then these initial conditions lead to heads as a result. If they fall within the purple area, then tails is the final outcome. From the fact that the initial conditions are uniformly distributed over I, we can conclude that the probability of heads, P_H, is simply equal to the proportion of the yellow area to the area of the interval I, that is,

$$P_H = \mathbb{P}((u, \omega) \in I \cap H) = \frac{\text{area of } I \cap H}{\text{area of } I}. \tag{16}$$

Observe that this is approximately equal to one-half. Note that as the square I moves to the top right corner of the graph, the stripes become narrower and the gray and white zones alternate more quickly. We see that P_H comes closer to one half. Therefore, one might expect that $P_H \to \frac{1}{2}$ as $a, b \to \infty$. This is the idea behind the method of arbitrary functions. In Section 2.4 we will generalize this idea by taking any continuous initial distribution, and not just a uniform distribution over some interval I. Conversely, if I shrinks and moves to the bottom left corner of the graph, I becomes primarily white and therefore P_H approaches 1. At first glance this might sound strange, but it is coherent with reality when you think about it. Suppose the coin is tossed with a sufficiently low vertical and angular velocity such that heads faces up during all of the coin toss, then the chance of the coin landing tails is 0. In other words, $P_H = 1$ if u and ω are contained in an interval I which is sufficiently scrunched in the lower

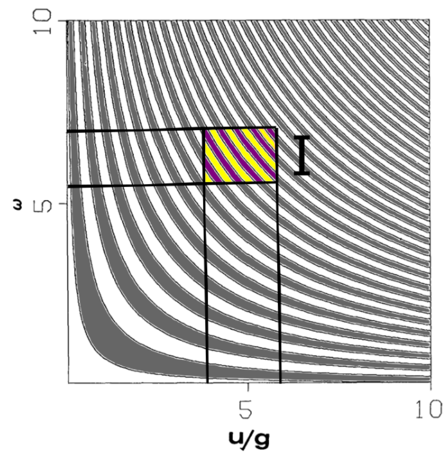

Figure 4: The yellow zones add up to $I \cap H$ and the purple zones add up to $I \cap T$.

left corner. Generally, if the interval I is shrunken enough, then it turns either primarily white or primarily gray, regardless of where in the graph I is located. So there could exist some kind of masterful magician who is able to 'defy the odds' and lands heads 100 times successively. Of course, he is not defying the odds, he is 'merely' controlling the way the coin leaves his hand very skillfully. That is, he is making sure the coin leaves his hand with certain initial values such that these are contained in an interval I such that $I \subseteq H$. It turns out that our own capacities determine the probability of heads. And if our capacities are skillful enough, one might be able to eliminate probability altogether. But most of us, if not all, have not mastered this kind of magicianship. Therefore, the coin toss remains an unpredictable affair.

2.4 The method of arbitrary functions

In Section 2.3 we have observed that when the initial conditions are uniformly distributed over some interval

$$I = \{(u, \omega) : a \leq u \leq a + 1, b \leq \omega \leq b + 1\}, \quad (17)$$

then we expect something like $P_H \to \frac{1}{2}$ as $a, b \to \infty$ to happen. This statement would only mean something if we could justify the assumption that the initial values are distributed in such a way. But it is quite absurd to assume that this distribution is uniform. And to go even further, it might be far fetched to assume that we have any knowledge of the ways in which the initial conditions are distributed (but more on that in Chapter 4). Since it is difficult to make a definitive statement about the natural occurrence of the initial distribution, it would be favorable if we could eliminate the need for such knowledge altogether. As this section will demonstrate, we will not be completely able to do such a thing, but only partly. The main theorem of this section, Theorem 2.2, shows us that P_H approaches half for an arbitrary continuous initial distribution, hence the method of *arbitrary* functions.

Definition 2.1. *For $f : X \to \mathbb{R}$, the **support** of f is the set of all points $x \in X$ such that $f(x) \neq 0$.*
*A function f is said to have **compact support** if the closure of the support of f is a compact set.*

Suppose that $p : \mathbb{R}^2 \to [0, 1]$ is a continuous probability density such that for any $A \subseteq \mathbb{R}^2$ we have

$$\mathbb{P}\big((u, \omega) \in A\big) = \int\int_A p(u, \omega) du d\omega.$$

Then

$$P_H = \mathbb{P}\big((u, \omega) \in H\big) = \int\int_H p(u, \omega) du d\omega.$$

Theorem 2.2. *Let $p(u, \omega)$ be a continuous probability density with support in the region $u > 0$, $\omega > 0$, and let β be a fixed constant satisfying $0 \leq \beta \leq \pi/2$. Then:*

$$\lim_{U\to\infty} P_H = \lim_{U\to\infty} \int\int_H p(u - U\cos\beta, \omega - a^{-1}U\sin\beta) du d\omega = \frac{1}{2}. \tag{18}$$

Proof. We want to know the limiting value of P_H as the support of $p(u, \omega)$ shifts to infinity. Because $U \cos\beta \geq 0$ and $a^{-1}U\sin\beta \geq 0$ for all $0 \leq \beta \leq \pi/2$ we see that this shifting of support is exactly what happens when

$$\lim_{U\to\infty} p(u - U\cos\beta, \omega - a^{-1}U\sin\beta). \tag{19}$$

This proves the first equality of the theorem.
Now on to the second equation of the theorem. Recall that H is the pre-image of heads. H is precisely the union of all the white areas seen in Figure 3, that is:

$$H = \bigcup_{n\in\mathbb{N}} \big\{(u, \omega) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} : \omega \in \frac{\pi g}{2u}\big[2n - \frac{1}{2}, 2n + \frac{1}{2}\big]\big\}. \tag{20}$$

This leads to the following equation:

$$P_H = \int_{u\in\mathbb{R}_{>0}} \left(\sum_{n\in\mathbb{N}} \int_{(2n-\frac{1}{2})\frac{\pi g}{2u}}^{(2n+\frac{1}{2})\frac{\pi g}{2u}} p(u - U\cos\beta, \omega - a^{-1}U\sin\beta) d\omega\right) du. \tag{21}$$

By setting $\omega' = \omega - a^{-1}U\sin\beta$, integration by substitution gives us the following:

$$P_H = \int_{U\cos\beta}^{\infty} \left(\sum_{n\in\mathbb{N}} \int_{(2n-\frac{1}{2})\frac{\pi g}{2u} - a^{-1}U\sin\beta}^{(2n+\frac{1}{2})\frac{\pi g}{2u} - a^{-1}U\sin\beta} p(u - U\cos\beta, \omega') d\omega'\right) du \tag{22}$$

The fact that the lower limit of u is equal to $U\cos\beta$ is due to the support of p; $p(u - U\cos\beta, \omega') = 0$ for all $u \leq U\cos\beta$.
As U goes to infinity, so does $U\cos\beta$ provided that $\beta < \frac{\pi}{2}$. If U goes to infinity then the lower bound of the range of the integral over u also goes to infinity. The length of the range of each integral over ω is $\frac{\pi g}{2u}$ and therefore tends to 0 as $U \to \infty$. Note that the point $\frac{2n\pi g}{2u} - a^{-1}U\sin\beta$ is the midpoint of the n^{th} interval for all $u > 0$. Using the continuity of p, each integral over ω' approaches the integrand evaluated at the midpoint multiplied by the length of the interval as $U \to \infty$. Now we implement this newly found equation into (22), so that P_H approaches

$$\int_{U\cos\beta}^{\infty}\left(\sum_{n\in\mathbb{N}} p(u - U\cos\beta, \frac{2n\pi g}{2u} - a^{-1}U\sin\beta)\frac{\pi g}{2u}\right)du, \tag{23}$$

as $U \to \infty$. Define $\Delta\omega = \frac{\pi g}{2u}$, then the sum inside the integral can be rewritten as

$$\sum_{n=1}^{\infty} p(u - U\cos\beta, 2n\Delta\omega - a^{-1}U\sin\beta)\Delta\omega. \tag{$*$}$$

Define A_ω in the following manner:

$$A_\omega = \sup\{u > 0 : p(u' - U\cos\beta, \omega - a^{-1}U\sin\beta) = 0 \text{ for all } u' \leq u\}. \tag{24}$$

As $U \to \infty$, also $A_\omega \to \infty$. Recalling that $\Delta\omega$ is proportional to u^{-1} we find that the values of $\Delta\omega$ for which ($*$) is nonzero become infinitely small as U approaches infinity. Note that

$$\begin{aligned}&\sum_{n\in\mathbb{N}} \inf_{2n\Delta\omega\leq\omega\leq(2n+1)\Delta\omega} p(u - U\cos\beta, \omega - a^{-1}U\sin\beta)\Delta\omega \leq (*)\\ &\leq \sum_{n\in\mathbb{N}} \sup_{2n\Delta\omega\leq\omega\leq(2n+1)\Delta\omega} p(u - U\cos\beta, \omega - a^{-1}U\sin\beta)\Delta\omega,\end{aligned} \tag{25}$$

and that, as $U \to \infty$, the first sum converges to half the lower Riemann sum and the last sum converges to half the upper Riemann sum. If p has compact support then it is Riemann integrable since it is continuous by assumption. So, we find that if p satisfies compact support then ($*$) converges to half the Riemann integral. Implementing this in (23) gives us that P_H approaches

$$\int_{U\cos\beta}^{\infty}\left(\frac{1}{2}\int_0^{\infty} p(u - U\cos\beta, \omega')d\omega'\right)du. \tag{26}$$

If p does not have compact support, it can be approximated with a sequence of functions with compact support and doing so leads to the same result.
Setting $u' = u - U\cos\beta$ and applying integration by substitution to (26) gives us

$$P_H \to \frac{1}{2}\int_0^{\infty}\int_0^{\infty}(u', \omega')du'd\omega' = \frac{1}{2} \quad \text{as } U \to \infty, \tag{27}$$

where the integral is equal to 1 since p is a density function. Therefore, the desired result is obtained. □

2.5 Some notes

Theorem 2.2 shows us that the probability of heads becomes independent of the initial distribution as long as it is continuous. Thus, if we can agree on the fact that this initial distribution is continuous, then we can make a definitive statement about the limiting value of P_H. So we must ask ourselves whether the occurrence of a discontinuous initial distribution is natural. Suppose that someone flips a coin with an initial distribution that is discontinuous at some point $\bar{u} \in \mathbb{R}_{\geq 0}$. Then, this would mean that this person is suddenly significantly less or more likely to throw a coin with the exact and very specific vertical speed $\bar{u}$. This seems highly unlikely to happen since it would imply that this person somehow has the capacity of recognizing this exact value. Therefore, based on reason, we could eliminate the occurrence of such a discontinuous initial distribution. This would render the limiting value of P_H to be independent of any *actual* initial distribution.

However, in real life, initial conditions do not exceed certain values. Therefore, an initial distribution with support at infinite values is merely a hypothetical situation. If we take the limiting value of its support then, surely, any continuous initial distribution does the job. But when the initial distribution is fixed, as it is in real life, this is no longer the case. In order to illustrate this, let us construct an initial density function f such that it is continuous and its support is confined in some finite interval, but the probability of heads is not nearly equal to one-half. Define real numbers $a_1, a_2, b_1, b_2 \in \mathbb{R}_{>0}$ such that

$$\{(u, \omega) : f(u, \omega) \neq 0\} \cap \{(u, \omega) : a_1 \leq u \leq a_2, b_1 \leq \omega \leq b_2\}^{\mathsf{c}} = \emptyset. \tag{28}$$

Now construct f in such a way as is shown in Figure 5. Here, the yellow area is equal to the intersection of H with $[a_1, a_2] \times [b_1, b_2]$. Observe that

$$\int_{H \cap [a_1, a_2] \times [b_1, b_2]} f(u, \omega) du d\omega >> \int_{T \cap [a_1, a_2] \times [b_1, b_2]} f(u, \omega) du d\omega. \tag{29}$$

Therefore, the probability of heads is not nearly equal to half. The idea is that, since the support of f is confined in some finite interval, the stripes which form the pre-images of heads and tails are not infinitely small. Therefore, we can construct a *continuous* function f which has high values for input values (u, ω) which lie in a yellow area and low values for (u, ω) which lie in a purple area. This causes P_H to deviate significantly from half. No matter how much the support of f is shifted towards infinity, as long as it is finite one can always construct such a continuous density function f. In fact, for any $p \in [0, 1]$ there exists a continuous density function such that it generates p as the probability of heads. This seems quite problematic. However, for an initial distribution to generate a probability that *significantly* deviates from $\frac{1}{2}$, it must peak at certain specific values (see Figure 5); it must possess certain unconventional or *eccentric* features. The actual occurrence of such an eccentric initial distribution is debatable and will be discussed in more detail in Chapter 4, specifically in Section 4.6.

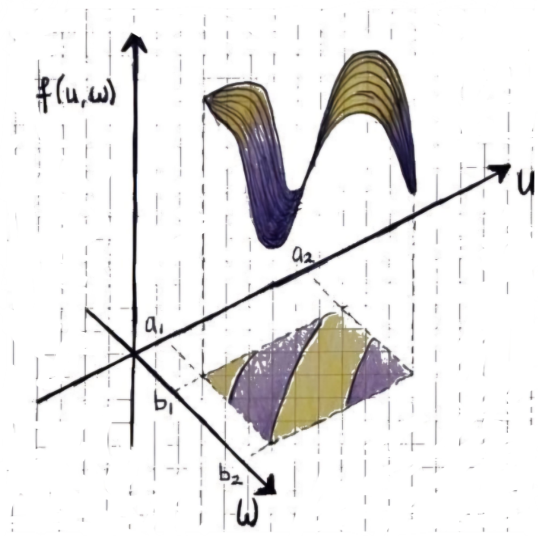

Figure 5: Construction of f.

Exclusion of the occurrence of these eccentric distributions only excludes those that generate a probability which *significantly* deviate from half. Most of the remaining, 'plausible', distributions still generate a probability whose value is close, but not equal, to half. We seem to be left with an interval of probabilities who are all likely to be 'the one', presuming that there is one true probability. We will also briefly touch on this question in Chapter 4, specifically in Section 4.5.

So, one might wonder how the method is applicable in real life, where infinitely large initial conditions do not exist? The answer is simple; it is not. However, it does show us that as some variable grows bigger and bigger, then the amount of continuous density functions that generate a significantly deviating probability grows smaller and smaller. One could say that this has inspired the interpretation of probability discussed in Section 4.5.

2.6 Other games of chance

The method of arbitrary functions is in principle applicable to any random phenomenon of deterministic nature. The main examples of applications so far lie within games of chance. We briefly discuss some of these to strengthen our grasp on the material.

2.6.1 The wheel

Assume a symmetrically balanced wheel on which equally sized alternating stripes of gray and white are painted and which is provided with a fixed pointer

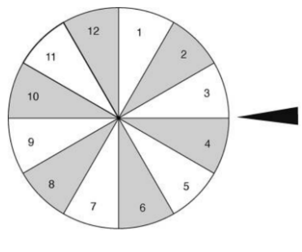

Figure 6: The wheel ((3), p.665).

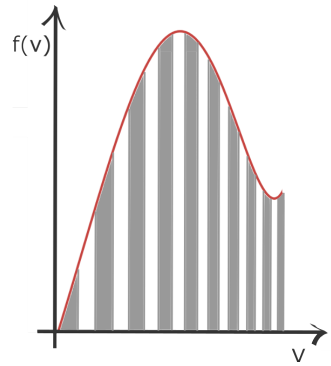

Figure 7: The density function f over the initial velocity.

(see Figure 6). When the wheel comes to rest, the fixed pointer points to white or gray, which is the outcome of the experiment. The outcome merely depends on one initial value, namely the velocity with which the wheel is initially spun. The probability that the fixed pointer reveals gray approaches one-half if either of the following two variables approaches infinity:

1. The number of painted stripes on the wheel,
2. The initial velocity.

We will not give a proof of this claim; hopefully, explaining it in words will provide enough clarity on the matter. Take a look at Figure 7. Here, f demonstrates a continuous probability density function of the initial velocity v. Let P_G be the probability that gray is the final outcome. A gray bar in Figure 7 is the area below f over an interval of initial velocities that all lead to gray as outcome. Therefore, the total gray area is equal to P_G. Observing that the white and gray areas are approximately equal, this implies that P_G is approximately equal to half. Suppose that we could make these bars thinner somehow, then P_G would be even closer to one half. In order to thin out these bars, we could apply either of the two techniques enumerated above. Taking either of these two limits has the effect that the bars grow infinitely small, therefore $P_G \to \frac{1}{2}$ for any continuous density function f.

There is another technique to thin out the bars, but this would require a frictionless wheel. If the friction is equal to 0, then the wheel will continue spinning until eternity. Suppose that one takes a look at the wheel at time t and writes down the color the fixed pointer is pointing at. Hopf shows that the probability of it pointing at gray approaches $\frac{1}{2}$ as $t \to \infty$. In his book (5), Hopf develops the method of arbitrary functions limited to conservative mechanisms. These are mechanisms which are never at rest and one can therefore take the limiting

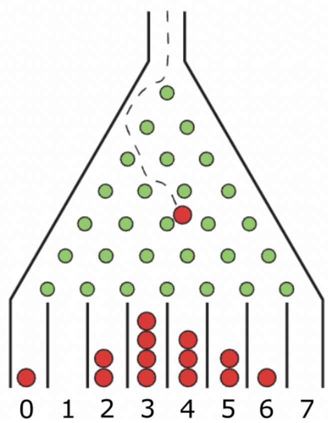

Figure 8: A 6-Galton board.

value of t to infinity. In this thesis, we will not concern ourselves with Hopf's studies on conservative mechanisms.

2.6.2 Galton board

Figure 8 shows a Galton board. The main idea behind this board is that a little ball is released into the Galton board from the top. This ball then makes its way down to one of the containers at the bottom. First, the ball touches the peg in the first row, depending on the manner in which this ball touches this peg, it either "bounces" off to the right or to the left, subsequently touching either the right of left peg in the second row.

Let us first assume that at each peg the ball is equally likely to bounce off to the right or left. What then is the probability of a single ball eventually placing itself in container i? Assume an n-Galton board. This board has $n+1$ containers and the number of pegs the ball touches is equal to n. The total number of ways one ball can travel through the board is equal to 2^n, because for every peg the ball touches, it has 2 options; to bounce off to the right or to the left. By assumption, these options are equally likely and therefore each of these 2^n ways of going down is as likely to happen as any other way. For the ball to end up in the i^{th} container, the ball must bounce to the right i times and $n-i$ times to the left. There are $\binom{n}{i}$ ways in which this happens. Therefore, the probability that a ball ends up in the i^{th} container is equal to

$$\frac{\binom{n}{i}}{2^n}. \tag{30}$$

However, in real life, the ball is not *exactly* as likely to bounce to the right as it is to bounce to the left for each peg. Were the tube as wide as in Figure 8, then we could relatively easily make sure that the ball touches the first peg on the left, making sure the ball never ends up at container 7. More generally, it is definitely plausible that the person who throws several balls through this tube has the tendency, intentionally or unintentionally, to throw the ball through the tube in such a way that it often ends up at the left of the first peg. Consequently, the probability that the ball ends up in container 7 is significantly lower than the one prescribed by (30). So, we see that the probability of a certain outcome is not completely independent of the initial distribution. In order to eliminate this dependency, we increase the length of the tube. Suppose that the length of the tube, l, is much larger; then the ball travels through the tube much longer and engages in much more bouncing against its walls. Therefore, much smaller differences in the way someone throws the ball into the tube lead to different containers in which the ball ends up. This means that an initial distribution has to be much more unconventional in order to generate a deviating probability. Therefore, it is much less likely that an initial distribution generates a probability that is significantly different from (30). But, in order to completely eliminate this dependency, one must take the limit of the length of the tube to infinity. For any continuous initial distribution we find that:

$$\mathbb{P}(\text{ball falls in container } i) \to \frac{\binom{n}{i}}{2^{n-1}} \text{ as } l \to \infty.$$

3 General Theory

The whole idea behind the method of arbitrary functions is that in the limit the final distribution is independent of the initial distribution as long as this initial distribution belongs to a certain class of functions. For the coin toss, the probability of heads and tails (the final distribution) depends on the distribution of the initial vertical and angular speed (the initial distribution), as long as it belongs to the class of continuous functions. In his article (12), von Plato discusses the various classes of functions for which the method works. For example, he discusses Poincaré's study of the roulette wheel, where Poincaré showed that if we assume that the arbitrary function has a derivative f' so that for some constant M it holds that $|f'| < M$, it follows that the probabilities of red and black are approximately equal.

However, it would be practical to develop a theory that is not just applicable to one particular game of chance. Hopf provided such a *general* theory. Engel unified and extended Hopf's findings in his book (4). He includes Hopf's work on dissipative mechanisms experiencing low levels of friction. The contents of this chapter, Section 3.2 in particular, are mainly based on Engel's book, along with some references to Billingsley's book (2).

In the preface of his book, Engel (1992) writes that "most applications of the method of arbitrary functions follow from the fact that the fractional part of the product of a real number t and an absolutely continuous random vector X converges to a distribution that is uniform on the unit hypercube as t tends to infinity". Section 3.2 therefore culminates in proving this statement, which is set out in Corollary 3.15. Its connection to the method of arbitrary functions might not be immediately apparent and will therefore be clarified in Section 3.1 with the help of the coin toss.

However, Engel goes a little further and also studies the convergence rate in order to determine the practical relevance of the method of arbitrary functions for specific examples. We will briefly touch on the topic of convergence rates in Section 3.3, but we will not discuss the relevant proofs. We will also not go into the higher dimensional cases. This is partly due to lack of time, and partly due to the similarities it shares with the one dimensional case[1].

In summary, in this section we will first use the coin toss once again to illustrate why we prove what we are about to prove. After that, in Section 3.2, we will develop the theory for one dimension and briefly mention the main theorem in relation to higher dimensions. Finally, in Section 3.3, we will discuss some other examples that demonstrate the generality of the theory.

3.1 The coin toss, again

As mentioned above, Engel writes that "most applications of the method of arbitrary functions follow from the fact that the fractional part of the product of a real number t and an absolutely continuous random vector X converges to

[1] I would like to highlight the fact that the contents of Engel's book considerably exceed the contents of this thesis, so, when interested, one should take a further look at his book (4).

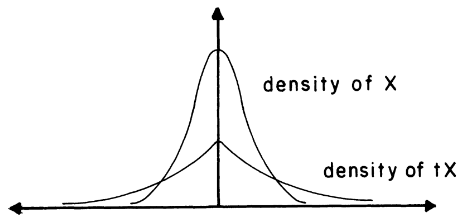

Figure 9: tX approaches a uniform distribution on $[0, 1]$ as $t \to \infty$ ((4), p.3).

a distribution that is uniform on the unit hypercube as t tends to infinity". Let us review what exactly this means. Let X be an absolutely continuous random vector and let $t \in \mathbb{R}$. Engel refers to $(tX)(\text{mod } 1)$ when he speaks of "the fractional part of the product of a real number t and an absolutely continuous random vector X". So, the essence of the statement is the following:

$$(tX)(\text{mod } 1) \to U_n \text{ as } t \to \infty, \tag{31}$$

where U_n is the uniform distribution on the unit hypercube $[0, 1]^n$. Looking at Figure 9, this statement is quite intuitive. However, actually proving it requires some work and makes up the entire focus of Section 3.2. But why do we put so much effort into proving this statement? What is its connection to the method of arbitrary functions? We illustrate this with the help of the coin toss.

Recall that H is the pre-image of heads and was equal to

$$H := \left\{(u, \omega) : \omega \in \bigcup_{u \in \mathbb{R}_{>0}} \bigcup_{n \in \mathbb{N}} \left[\Big(2n - \frac{1}{2}\Big)\frac{\pi g}{2u}, \Big(2n + \frac{1}{2}\Big)\frac{\pi g}{2u}\right]\right\}. \tag{32}$$

Define the following set:

$$A := \left\{(u, \omega) : \left(\frac{\omega u}{\pi g} + \frac{1}{4}\right)(\text{mod } 1) \leq \frac{1}{2}\right\}. \tag{33}$$

We claim that $H = A$.

Proof.

$$\begin{aligned}
(u, \omega) \in A &\iff \left(\frac{\omega u}{\pi g} + \frac{1}{4}\right)(\text{mod } 1) \leq \frac{1}{2} \\
&\iff \exists_{n \in \mathbb{N}} \text{ s.t. } \frac{\omega u}{\pi g} + \frac{1}{4} \in \Big[n, n + \frac{1}{2}\Big] \\
&\iff \exists_{n \in \mathbb{N}} \text{ s.t. } \omega \in \left[\Big(2n - \frac{1}{2}\Big)\frac{\pi g}{2u}, \Big(2n + \frac{1}{2}\Big)\frac{\pi g}{2u}\right] \\
&\iff (u, \omega) \in H.
\end{aligned}$$

□

Suppose that the initial angular speed ω is distributed by some absolutely continuous distribution function X. Then

$$\begin{aligned} P_H(u) &= \mathbb{P}\big((u, X) \in H\big) \\ &= \mathbb{P}\big((u, X) \in A\big) \\ &= \mathbb{P}\bigg(\Big(\frac{uX}{\pi g} + \frac{1}{4}\Big)(\text{mod } 1) \leq \frac{1}{2}\bigg). \end{aligned} \tag{34}$$

Combining (31) and (34) allows us to arrive at the desired conclusion that $P_H(u)$ approaches $\frac{1}{2}$ as $u \to \infty$.

3.2 The one-dimensional case

As mentioned before, this entire section has one goal; that of proving (31). More specifically, we will prove (31) limited to one-dimensional random vectors X:

$$(tX)(\text{mod } 1) \to U \text{ as } t \to \infty, \tag{35}$$

where U is the uniform distribution over $[0, 1]$. At the end of this section we will briefly cover the higher dimensional case.

When we refer to a random variable X in this subsection, we mean a one-dimensional real-valued random variable.

Definition 3.1. *Let $X, X_1, X_2, \dots$ be a sequence of random variables with distribution functions $F(x), F_1(x), F_2(x), \dots$. If $F_k(x)$ converges to $F(x)$ at every point of continuity of F, as k tends to infinity, then we say that $X_1, X_2, \dots$ converges in the* weak-star topology *to X.*

Definition 3.2. *Let X be a random variable. The* characteristic function *of X, denoted by $\hat{f}$, is defined by*

$$\hat{f}(t) = \mathbb{E}[e^{itX}].$$

Furthermore, the values of the characteristic function at $2\pi m$, $m \in \mathbb{Z}$, are called the Fourier coefficients *of X.*

Definition 3.3. *A probability space is a measure space $(\Omega, \mathcal{A}, \mu)$ such that $\mu(\Omega) = 1$.*

Definition 3.4. *Let X be a random variable. The measure μ_X attached to X is given by $\mu_X : \mathcal{B}(\mathbb{R}) \to [0, 1], A \mapsto \mathbb{P}(X \in A)$, where $\mathcal{B}(\mathbb{R})$ is the Borel algebra.*

Proposition 3.5. *Let X be a random variable. Then μ_X is a probability measure and hence $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ is a probability space.*

Proof. The Borel algebra $\mathcal{B}(\mathbb{R})$ is a σ-algebra by definition. So what's left to prove is:

(i) $\mu_X(\emptyset) = 0$;

(ii) $\mu_X(\mathbb{R}) = 1$;

(iii) For every sequence $\{A_i\}_{i\in\mathbb{N}}$ such that $A_i \in \mathcal{B}(\mathbb{R})$ for all i, and $A_i \cap A_j = \emptyset$ for all $i \neq j$, it holds that

$$\mu_X\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_X(A_i).$$

Because for all $A \in \mathcal{B}(\mathbb{R})$, $\mu_X(A)$ is defined as the probability of the event A, (i), (ii) and (iii) follow immediately from the axioms of probability. □

Definition 3.6. *Distribution functions F_n are said to* converge weakly *to F if $\lim_n F_n(x) = F(x)$ for every continuity point x of F. Notation: $F_n \Rightarrow F$.*

Suppose that $X, X_1, X_2, \ldots$ are random variables with distribution functions $F, F_2, F_3, \ldots$. Then X_n converges in the weak-star topology to X iff $F_n \Rightarrow F$. This is simply a matter of combining Definitions 3.1 and 3.6.

Definition 3.7. *Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures. Define $\mu_n \Rightarrow \mu$ iff $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all A of the form $A = (-\infty, x]$ such that $\mu\{x\} = 0$.*

Proposition 3.8. *Let $X, X_1, X_2, \ldots$ be random variables with corresponding distribution functions $F, F_1, F_2, \ldots$. Then $F_n \Rightarrow F$ iff $\mu_n \Rightarrow \mu$.*

Proof. Let x be an arbitrary point. Then:

$$\begin{aligned}\mu_X\{x\} &= \mathbb{P}(X = x)\\ &= \mathbb{P}(X \leq x) - \mathbb{P}(X < x)\\ &= \lim_{t\downarrow x} F(t) - \lim_{t\uparrow x} F(t).\end{aligned}$$

From this, it follows that $\mu_X\{x\} = 0$ if and only if F is continuous at x. This statement combined with the fact that

$$\mu_X(-\infty, x] = F(x) \quad \text{and} \quad \mu_{X_n}(-\infty, x] = F_n(x),$$

proves this proposition. □

Definition 3.9. *Let $(\Omega, \mathcal{F}, P)$ be a probability measure space. The* support *of P is the smallest closed set $A \in \mathcal{F}$ such that $P(A) = 1$.*

Assume a probability measure μ such that its support it contained in $[0, 1]$. Define its Fourier coefficient by

$$c_k(\mu) = \int_0^1 e^{2\pi i k x} d\mu(x). \tag{36}$$

Following Billingsley, we will show that the measure μ can be recovered from these Fourier coefficients[2]. Define the partial sum $s_l(t)$ and Césaro average $\sigma_m(t)$:

$$s_l(t) = \frac{1}{2\pi}\sum_{k=-l}^{l} c_k(\mu)e^{ikt}, \qquad \sigma_m(t) = \frac{1}{m}\sum_{l=0}^{m-1} s_l(t). \tag{37}$$

If we write $\sigma_m(t)$ in full length, we obtain

$$\sigma_m(t) = \frac{1}{2\pi m}\int_0^1 \left(\sum_{l=0}^{m-1}\sum_{k=-l}^{l} e^{2\pi ik(x-t)}\right)d\mu(x). \tag{38}$$

Consider the following trigonometric equality (whose proof can be found in the appendix in (2)):

$$\sum_{l=0}^{m-1}\sum_{k=-l}^{l} e^{ikx} = \frac{\sin^2(\frac{1}{2}mx)}{\sin^2(\frac{1}{2}x)}. \tag{39}$$

Now (38) can be rewritten as

$$\sigma_m(t) = \frac{1}{2\pi m}\int_0^1 \frac{\sin^2\left(\frac{1}{2}m2\pi(x-t)\right)}{\sin^2\left(\frac{1}{2}2\pi(x-t)\right)}d\mu(x). \tag{40}$$

Suppose that $0 < a < b < 1$ and take the integral of σ_m over (a, b). Applying Fubini's theorem then leads to:

$$\begin{aligned}
\int_a^b \sigma_m(t)dt &= \int_0^1 \left(\frac{1}{2\pi m}\int_a^b \frac{\sin^2\left(\frac{1}{2}m2\pi(x-t)\right)}{\sin^2\left(\frac{1}{2}2\pi(x-t)\right)}dt\right)d\mu(x) \\
&= \int_0^1 \left(\frac{1}{2\pi m}\int_a^b \frac{\sin^2\left(\frac{1}{2}m2\pi(t-x)\right)}{\sin^2\left(\frac{1}{2}2\pi(t-x)\right)}dt\right)d\mu(x) \\
&= \int_0^1 \left(\frac{1}{2\pi m}\int_{(a-x)2\pi}^{(b-x)2\pi} \frac{\sin^2(\frac{1}{2}ms)}{\sin^2(\frac{1}{2}s)}ds\right)d\mu(x),
\end{aligned} \tag{41}$$

where the second equality is due to the fact that $\sin^2(x) = \sin^2(-x)$. Furthermore, for all $0 < \delta < \pi$ the following holds:

$$\frac{1}{2\pi m}\int_{\delta<|s|<\pi} \frac{\sin^2(\frac{1}{2}ms)}{\sin^2(\frac{1}{2}s)}ds \to 0 \quad \text{and} \quad \frac{1}{2\pi m}\int_{|s|<\delta} \frac{\sin^2(\frac{1}{2}ms)}{\sin^2(\frac{1}{2}s)}ds \to 1, \tag{42}$$

as $m \to \infty$. Define:

$$(*) = \frac{1}{2\pi m}\int_{(a-x)2\pi}^{(b-x)2\pi} \frac{\sin^2(\frac{1}{2}ms)}{\sin^2(\frac{1}{2}s)}ds. \tag{43}$$

If $0 < x < a$, then $(b-x)2\pi > (a-x)2\pi > 0$. Therefore, $(*) \to 0$ as $m \to \infty$. Similarly, if $b < x < 1$ then $(*) \to 0$. If $a < x < b$, then $(b-x)2\pi > 0 > (a-x)2\pi$

[2] For more details on this, I refer to pages 351-352 in (2).

and therefore we obtain $(*) \to 1$. Combining all these results, we obtain the following:

$$\lim_m \int_a^b \sigma_m(t)dt = \int_a^b d\mu(x) = \mu(a,b]. \tag{44}$$

Finally, applying the bounded convergence theorem gives us

$$\mu(a,b] = \int_a^b \lim_m \sigma_m(t)dt, \tag{45}$$

provided that $\mu\{a\} = \mu\{b\} = 0$ and $0 < a < b < 1$.

Since the Césaro averages σ_m are determined by the Fourier coefficients, we see that μ can be recovered from these coefficients. On page 352, Billingsley uses (45) to make the following statement:

> *Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures whose support is contained in $[0,1]$. If*
>
> $$\lim_{n\to\infty} c_m(\mu_n) = c_m(\mu)$$
>
> *for all $m \in \mathbb{Z}$ and if $\mu\{0\} = \mu\{1\} = 0$, then $\mu_n \Rightarrow \mu$.*

We are now in a position to prove the following proposition.

Proposition 3.10. *Suppose that $X_1, X_2, \ldots$ is a sequence of random variables with distribution functions $F_1, F_2, \ldots$ such that $F_i(0) = 0$ and $F_i(1) = 1$ for all i. Let X be a random variable with distribution function F such that $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 0$. Then $X_1, X_2, \ldots$ converges in the weak-star topology to X iff given any $m \in \mathbb{Z}$ the sequence of Fourier coefficients associated with m (as in Definition 3.2) converges to the corresponding Fourier coefficient of X.*

Proof. $\Leftarrow$: By assumption we know that

$$\mu_{X_i}[0,1] = \mathbb{P}[0 \le X_i \le 1] = F_i(1) - F_i(0) = 1$$

for all i. Therefore, the support of μ_{X_i} is contained in $[0,1]$. Furthermore, by assumption we obtain

$$\mu_X\{0\} = \mathbb{P}(X = 0) = 0 = \mathbb{P}(X = 1) = \mu_X\{1\}.$$

Since for any probability measure μ with characteristic function $\hat{f}$ supported by $[0,1]$ we have

$$c_m(\mu) = \int_0^1 e^{2\pi i m x} d\mu(x) = \int_{\mathbb{R}} e^{2\pi i m x} d\mu(x) = \hat{f}(2\pi i),$$

we find:

$$\lim_n c_m(\mu_{X_n}) = \lim_n \hat{f}_n(2\pi m) = \hat{f}(2\pi m) = c_m(\mu_X),$$

for all $m \in \mathbb{Z}$. Applying the statement preceding this proposition, we find that $\mu_{X_n} \Rightarrow \mu_X$. As we have seen, this is equivalent to $F_n \Rightarrow F$, which in turn is equivalent to the convergence of F_n to F in the weak-star topology.
$\Rightarrow$: Trivial. □

Lemma 3.11. *Assume X is a random variable with distribution function F and characteristic function $\hat{f}$. Let $\hat{f}_t$ denote the characteristic function of $(tX)(mod\ 1)$. Then for all $m \in \mathbb{Z}$:*

$$\hat{f}_t(2\pi m) = \hat{f}(2\pi mt).$$

Proof.

$$\begin{aligned}
\hat{f}_t(2\pi m) &= \mathbb{E}\Big[e^{i2\pi m\big((tX)(\text{mod } 1)\big)}\Big] \\
&= \int_{\mathbb{R}} e^{i2\pi m((tx)(\text{mod } 1))} d\mu_X(x) \\
&= \sum_{k\in\mathbb{Z}} \int_{[k,k+1)} e^{i2\pi m((tx)(\text{mod } 1))} d\mu_X(x) \\
&= \sum_{k\in\mathbb{Z}} \int_{[k,k+1)} e^{i2\pi m(tx-k)} d\mu_X(x) \\
&= \sum_{k\in\mathbb{Z}} \int_{[k,k+1)} e^{i2\pi mtx} d\mu_X(x) \\
&= \hat{f}(2\pi mt).
\end{aligned}$$

The second to last equality is due to the fact that $e^{i2\pi mx}$ is 1-periodic. □

Theorem 3.12 (Poincaré, Borel, Fréchet, Kemperman)**.** *Let X be a real valued random variable with characteristic function $\hat{f}(t)$. Denote by U a distribution uniform on the unit interval. Then $(tX)(mod\ 1)$ converges to U in the weak-star topology as t tends to infinity, iff $\lim_{|t|\to\infty} \hat{f}(t) = 0$.*

Proof. $\Rightarrow$: Suppose that (tX)(mod1) converges to U in the weak-star topology as $t \to \infty$. Observe that (tx)(mod1) $\in [0,1]$ for all $x \in \mathbb{R}$ and therefore it is true that the support of (tx)(mod1) is contained in $[0,1]$ for all t. Suppose that μ is the measure that corresponds with U, then $\mu\{0\} = \mu\{1\} = 0$. Denote the measure that corresponds with (tx)(mod1) $\in [0,1]$ by μ_t. By Proposition 3.10 we find that

$$c_m(\mu_t) \to c_m(\mu),$$

as $t \to \infty$ for each $m \in \mathbb{Z}$. Due to Lemma 3.11 we know that

$$c_m(\mu_t) = \int_0^1 e^{2\pi imx} d\mu_t(x) = \hat{f}_t(2\pi m) = \hat{f}(2\pi mt).$$

If we can now prove that $c_m(\mu)$ is zero for all m, then we find that

$$\lim_{|t|\to\infty} \hat{f}(2\pi mt) = \lim_{|t|\to\infty} c_m(\mu_t) = c_m(\mu) = 0,$$

for any real value t. And therefore, $\lim_{|t|\to\infty} \hat{f}(t) = 0$. So let us prove that $c_m(\mu) = 0$. Let $m \in \mathbb{Z}$ be arbitrary. The density function of the uniform distribution over interval $[0,1]$ is equal to $g(t) = \mathbb{1}_{[0,1]}$. So:

$$\begin{aligned} c_m(\mu) &= \mathbb{E}[e^{2\pi i m U}] \\ &= \int_0^1 e^{2\pi i m t} dt = \frac{1}{2\pi i m}(e^{2\pi i m} - 1) \\ &= \frac{1}{2\pi i m}(-1 + \cos(2\pi m) + i \sin(2\pi m)) = 0. \end{aligned}$$

$\Leftarrow$: Suppose that $\lim_{|t|\to\infty} \hat{f}(t) = 0$. Then also:

$$c_m((tX)(\text{mod}1)) = \hat{f}(2\pi m t) \to 0 \text{ as } t \to \infty.$$

By Proposition 3.10 we can now conclude that $(tX)(\text{mod}1) \to U$ in the weak-star topology as $t \to \infty$. □

Definition 3.13. *Suppose that $(\Omega, \mathcal{A}, \mu)$ is a probability measure space and that δ is a nonnegative measurable function. Define a measure ν by*

$$\nu(A) = \int_A \delta d\mu,$$

where $A \in \mathcal{A}$. The measure ν is said to have density *δ with respect to μ.*

The following proposition can be found as Theorem 26.1 in Billingsley's book (2).

Proposition 3.14. *If μ has a density w.r.t. the Lebesgue measure, then $\hat{f}(t) \to 0$ as $|t| \to \infty$.*

Corollary 3.15. *If X is a random variable such that μ_X has a density w.r.t. the Lebesgue measure, then $(tX)(\text{mod } 1)$ converges in the weak-star topology to a distribution uniform on the unit interval.*

Proof. Follows directly from Proposition 3.14 and Theorem 3.12. □

But, what, beyond the above definition, does it exactly mean for a measure μ to have a density w.r.t. the Lebesgue measure? By the Radon-Nikodym theorem, if there exists a measurable function $g : \mathbb{R} \to [0, \infty]$ such that

$$\mu(A) = \int_A g(x) d\lambda(x), \tag{46}$$

for all $A \in \mathcal{B}(\mathbb{R})$, then we say that μ is absolutely continuous with respect to the Lebesgue measure λ, denoted by $\mu \ll \lambda$. We know that $\mu \ll \lambda$ iff for each $\epsilon > 0$ there exists $\delta > 0$ such that each $A \in \mathcal{B}(\mathbb{R})$ that satisfies $\lambda(A) < \delta$ also satisfies $\mu(A) < \epsilon$. So, let X be a random variable supported by $[0,1]$, then μ_X has a density w.r.t. the Lebesgue measure iff $\mu_X \ll \lambda$.

Another condition under which μ_X has a density w.r.t. the Lebesgue measure is absolute continuity of the corresponding distribution function. Assume that X is a random variable supported by $[0, 1]$ with a distribution function F which is absolutely continuous. Absolute continuity of a real valued function on some compact interval implies the existence of a Riemann integrable function f on $[0, 1]$ such that

$$F(b) - F(a) = \int_a^b f(t)dt. \tag{47}$$

Now, if f is a bounded Riemann-integrable function on $[0, 1]$, then f is Lebesgue integrable and the Riemann and Lebesgue integrals of f coincide. This gives us

$$\mu_X(a, b] = F(b) - F(a) = \int_a^b f(t)dt = \int_a^b f(t)d\lambda(t). \tag{48}$$

Therefore, we can conclude that μ_X has a density w.r.t. the Lebesgue measure.

Before we proceed with some applications, we briefly take a look at Theorem 3.16 which concerns the higher dimensions. When we speak of an n-dimensional random vector X, we mean a vector $X = (X_1, X_2, ..., X_n)$ where X_i is a real valued random variable for each $1 \leq i \leq n$. Set $(tX)(\text{mod } 1) = ((tX_1)(\text{mod } 1), ..., (tX_n)(\text{mod } 1))$. Let U_n be the uniform distribution over $[0, 1]^n$.

Theorem 3.16. *Let X be an n-dimensional random vector with density w.r.t. the Lebesgue measure, then $(tX_n)(mod\ 1)$ converges in the weak-star topology to U_n as $t \to \infty$.*

3.3 Examples

We have already seen how Corollary 3.15 applies to the coin toss. Does it also apply to the roulette wheel and Galton board, discussed in Section 2.6? Its applicability to the roulette wheel we will be made exact in this section. However, the dynamical properties of the Galton board are too complicated and will be dismissed. To compensate for this absence, we will discuss two other games of chance, just for fun[3].

3.3.1 The wheel

Recall the wheel from Section 2.6, shown again in Figure 10. Someone spins the wheel, and the color the red arrow points to when the wheel comes to rest is the outcome of the experiment. The outcome, white or gray, is determined by the initial angular velocity. Let n be the number of white and gray stripes on the wheel. Furthermore, let $\mu > 0$ be the coefficient of friction and assume that it is constant. Then

$$\frac{d^2\theta(t)}{dt^2} = -\mu. \tag{49}$$

[3]These two examples, set out in Section 3.3.2 and 3.3.3, appear quite out of the blue. Their place in this thesis exists purely because I enjoy them and thought someone else might too. Therefore, these two sections can easily be skipped and do not contribute to the structure of the thesis.

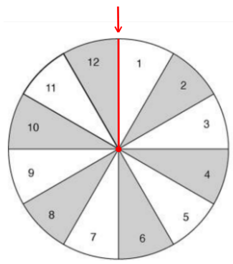

Figure 10: The wheel ($n = 12$).

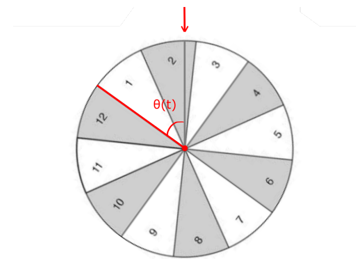

Figure 11: The angular position of the wheel at time t, denoted by $\theta(t)$.

Since $\theta(0) = 0$, we find that

$$\theta(t) = -\frac{1}{2}\mu t^2 + \omega t, \tag{50}$$

where ω is the initial angular velocity. Assume that the wheel comes to rest at time t_0, then

$$0 = \frac{d\theta(t_0)}{dt} = -\mu t_0 + \omega. \tag{51}$$

Therefore,

$$\theta(t_0) = \theta\left(\frac{\omega}{\mu}\right) = \frac{\omega^2}{2\mu}. \tag{52}$$

This allows us to conclude that if

$$\frac{\omega^2}{2\mu} \in \bigcup_{k\in\mathbb{N}} \left[\frac{2\pi}{n}2k, \frac{2\pi}{n}(2k+1)\right], \tag{53}$$

then white is the outcome. Because

$$\begin{aligned} &\exists_{k\in\mathbb{N}} \text{ s.t. } \frac{\omega^2}{2\mu} \in \left[\frac{2\pi}{n}2k, \frac{2\pi}{n}(2k+1)\right] \\ &\iff \exists_{k\in\mathbb{N}} \text{ s.t. } \frac{n\omega^2}{8\mu\pi} \in [k, k+1] \\ &\iff \frac{n\omega^2}{8\mu\pi} \bmod 1 \leq \frac{1}{2}, \end{aligned} \tag{54}$$

we find that

$$\mathbb{P}(\text{white}) = \mathbb{P}\left(\frac{n\omega^2}{8\mu\pi} \bmod 1 \leq \frac{1}{2}\right), \tag{55}$$

where $\mu > 0$ is a constant. Now, Corollary 3.15 allows us to conclude that

$$\mathbb{P}\left(\frac{n\omega^2}{8\mu\pi} \bmod 1 \leq \frac{1}{2}\right) \to \mathbb{P}\left(U \leq \frac{1}{2}\right) = \frac{1}{2} \quad \text{as } n \to \infty,$$

if the initial angular velocity ω has an absolutely continuous distribution function.

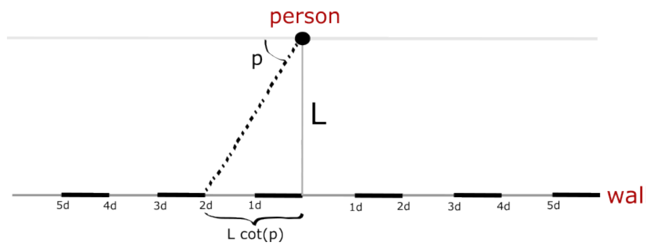

Figure 12: A bird's eye view of the dart experiment.

3.3.2 Throwing of darts

Consider throwing a dart at a wall. The wall is painted with black and white stripes. Assume that when someone throws a dart at the wall, the dart travels its way to the wall in a line parallel to the ground. Note that this is not the original game "darts", but a simplified version. Fix the width of the stripes d and the distance between the person and the wall L. Now the only variable is the angle p, see Figure 12. It should be clear that when the person stands very close to the wall or if the width of the stripes is large, then it is easy to aim at black. In this case, there is little randomness in the entire experiment. However, if d is small or L large, the result becomes more difficult to manipulate. Therefore, we expect that when $\frac{L}{d} \to \infty$ it follows that $\mathbb{P}(\text{black}) \to \frac{1}{2}$. Let us make this exact using the results from Section 3.2.

Suppose that $0 < p \leq \frac{\pi}{2}$. The distance between the middle point and the place on the wall where the dart ends up is equal to

$$\tan(\frac{\pi}{2} - p)L = \cot(p)L;$$

see Figure 12. We find that if there exists an $n \in \mathbb{N}$ such that

$$2nd \leq \cot(p)L \leq (2n+1)d, \tag{56}$$

then it follows that the dart ends up in black. For $\frac{\pi}{2} \leq p < \pi$, this distance is equal to

$$\tan(p - \frac{\pi}{2})L = -\cot(p)L.$$

It follows that if there exists $n \in \mathbb{N}_{>0}$ such that

$$(2n-1)d \leq -\cot(p)L \leq 2nd, \tag{57}$$

then black is the outcome. Since $\cot(-p) = -\cot p$, we can conclude that

$$\begin{aligned} \mathbb{P}(\text{black}) &= \sum_{n\in\mathbb{Z}} \mathbb{P}\Big(2nd \leq \cot(p)L \leq (2n+1)d\Big) \\ &= \mathbb{P}\left(\left(\frac{\cot p}{2}\frac{L}{d}\right)(\text{mod } 1) \leq \frac{1}{2}\right). \end{aligned} \tag{58}$$

Assuming that $\cot p$ has an absolutely continuous distribution function, we gather by Corollary 3.15 that

$$\left(\frac{\cot p}{2}\frac{L}{d}\right)(\text{mod } 1) \to U \qquad \text{as } L/d \to \infty, \tag{59}$$

where U is the uniform distribution over $[0,1]$. From this it immediately follows that

$$\mathbb{P}(\text{black}) \to \frac{1}{2}, \qquad \text{as } L/d \to \infty. \tag{60}$$

As we mentioned at the beginning of this chapter, Engel (1992) studies the convergence rate as well. Simply copying his findings on p. 48-50 gives us the following result:

$$\left|\mathbb{P}(\text{black}) - \frac{1}{2}\right| \leq \frac{6d}{L}. \tag{61}$$

So, for example, if the thrower is 3 meters away from the wall and the width of the stripes is less than one fifth of a centimeter, then the probability of hitting black is within 0.001 of one half. However, no one would play this game, because what is the fun of a game if you cannot get good at it?

3.3.3 Billiards

Assume a frictionless square billiard table. So once the ball has been hit, it keeps hitting cushions and bouncing on the table forever. The relevant feature of the dynamics of billiards is that the angle of incidence is equal to the angle of reflection. The purpose of this subsection is to show that no matter at which position and in which direction the ball is hit, the ball has no preference to be positioned in a certain area of the table as long as enough time has passed. Or, mathematically expressed; as $t \to \infty$, the distribution of the position of the ball at time t is uniform over $[0,1]^2$. Denote the position of the billiard ball at time t by $(x_1(t), x_2(t))$. For simplicity, assume that the ball's initial position is $(0,0)$. See an example of a trajectory of a billiard ball in Figure 13. Denote the initial velocity by (v_1, v_2). Define $y(t) = (y_1(t), y_2(t))$ such that

$$y_1(t) = v_1 t, \qquad y_2(t) = v_2 t. \tag{62}$$

Looking at Figure 14, it hopefully becomes clear that $x_1(t)$ is equal to $y_1(t)(\text{mod } 1)$ or to $(2 - y_1(t))(\text{mod } 1)$. In particular,

$$x_1(t) = \begin{cases} (v_1 t)(\text{mod } 2), & \text{if } y_1(t)(\text{mod } 2) \leq 1 \\ 2 - (v_1 t)(\text{mod } 1), & \text{if } y_1(t)(\text{mod } 2) > 1. \end{cases} \tag{63}$$

Similarly, $x_2(t)$ is given by:

$$x_2(t) = \begin{cases} (v_2 t)(\text{mod } 2), & \text{if } y_2(t)(\text{mod } 2) \leq 1 \\ 2 - (v_2 t)(\text{mod } 1), & \text{if } y_2(t)(\text{mod } 2) > 1. \end{cases} \tag{64}$$

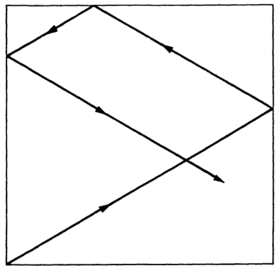

Figure 13: Trajectory of a billiard ball with initial position $(0,0)$, ((4), p.81).

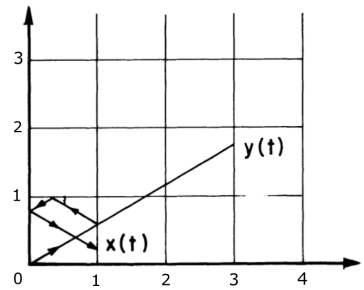

Figure 14: (4), p.82.

Define $\overline{y} = y \bmod 2$. Define the following four sets:

$$\begin{aligned} A_1 &= \{t : \overline{v_1 t} \le 1 \text{ and } \overline{v_2 t} \le 1\}; \\ A_2 &= \{t : \overline{v_1 t} \le 1 \text{ and } \overline{v_2 t} > 1\}; \\ A_3 &= \{t : \overline{v_1 t} > 1 \text{ and } \overline{v_2 t} \le 1\}; \\ A_4 &= \{t : \overline{v_1 t} > 1 \text{ and } \overline{v_2 t} > 1\}. \end{aligned} \tag{65}$$

Then:

$$(x_1(t), x_2(t)) = \begin{cases} (\overline{v_1 t}, \overline{v_2 t}), & \text{if } t \in A_1 \\ (\overline{v_1 t}, 2 - \overline{v_2 t}), & \text{if } t \in A_2 \\ (2 - \overline{v_1 t}, \overline{v_2 t}), & \text{if } t \in A_3 \\ (2 - \overline{v_1 t}, 2 - \overline{v_2 t}), & \text{if } t \in A_4. \end{cases} \tag{66}$$

Now assume that the 2-dimensional random variable (v_1, v_2) has an absolutely continuous distribution function. Then $(\overline{v_1 t}, \overline{v_2 t})$ converges in the weak-star topology to the uniform distribution over $[0,2]^2$. Let

$$I = (a,b) \times (a,b) \subseteq [0,1]^2.$$

Then:

$$\begin{aligned} \mathbb{P}(x(t) \in I) &= \sum_{i=1}^{4} \mathbb{P}(x(t) \in I \big| t \in A_i) \\ &= \mathbb{P}\big((\overline{v_1 t}, \overline{v_2 t}) \in I\big) \\ &+ \mathbb{P}\big((\overline{v_1 t}, \overline{v_2 t}) \in (a,b) \times (2-b, 2-a)\big) \\ &+ \mathbb{P}\big((\overline{v_1 t}, \overline{v_2 t}) \in (2-b, 2-a) \times (a,b)\big) \\ &+ \mathbb{P}\big((\overline{v_1 t}, \overline{v_2 t}) \in (2-b, 2-a) \times (2-b, 2-a)\big) \\ &\to \frac{(b-a)^2 + 2(2-a-(2-b))(b-a) + (2-a-(2-b))^2}{4} \\ &= (b-a)^2 \qquad \text{as } t \to \infty. \end{aligned} \tag{67}$$

So we see that $x(t)$ has the same distribution function as U_2, and is therefore uniformly distributed over $[0,1]^2$ as $t \to \infty$.

4 An alternative interpretation of probability

As briefly mentioned in Chapter 1, influential interpretations of probability are challenged by some serious shortcomings. This indicates that, up until now, there does not exist a sound and fruitful interpretation of probability. It is even unclear what criteria a good interpretation should meet. Since probability is everywhere around us, it is important to reflect on this debate and, perhaps, to find other ways of interpreting probability. Numerous people have created such an alternative interpretation with the use of the method of arbitrary functions, such as Rosenthal (10) and Strevens (11).

We will start this chapter by briefly going over some of the criteria a good interpretation of probability should meet. After that, in Section 4.2, we will discuss the most familiar interpretation of probability, frequentism, and how it fails some of the criteria. This hopefully demonstrates the importance of finding an alternative interpretation of probability. In Section 4.3 the potential of an interpretation inspired by the method of arbitrary functions is outlined. We will denote such an interpretation in question by *dynamical probability*. It is an umbrella term for all interpretations of probability whose basic idea is the same as that of the method of arbitrary functions; it is a kind of probability that is derived through the physical or dynamical properties of an experiment. In Section 4.4, we will discuss how dynamical probability is limited to a certain kind of probability and experiment. What the embodiment of dynamical probability should exactly be has been studied by numerous people, in particular Rosenthal (10) and Strevens (11). We will start by giving Rosenthal's definition in Section 4.5. Evidently, this definition runs into some issues, among them the problem of the initial distribution. This is the problem that the rest of this thesis focuses on, which is outlined in Section 4.6. We will see how Rosenthal and Strevens deal with this issue in Sections 4.6.1 and 4.6.2, respectively.

4.1 Criteria of a good interpretation

What exactly should we look for when we want to find a good interpretation of probability? As established in the introduction, there does not exist a particular list of criteria an interpretation should meet. According to Abrams (2010, p.346), this divergence may be healthy, at least in case that different interpretations of probability are appropriate for different contexts. The list of criteria composed below is mainly inspired by Abrams (1) and (8). Apart from the criteria listed below, an interpretation should of course also satisfy the basic criteria of a good definition, such as; non-circularity, preciseness, and clarity.

(1) *Ascertainability.* There must exist some kind of method which assigns values to the probabilities.

(2) *Satisfaction of probability axioms.* An interpretation ideally follows a certain calculus of probability. This means that the values an interpretation assigns to events follow certain rules or axioms. However, there exist numerous different calculi of probability. Some of these are variations on

Kolmogorov's calculus, and some of them are completely different. In order not to complicate things too much, we will omit this debate and take Kolmogorov's famous axioms as the probability calculus which a good interpretation should follow.

Definition 4.1 (Kolmogorov's probability calculus)**.** *Let Ω be a non-empty set. Let $\mathcal{F}$ be a field (or algebra) that is closed under complementation and union. Let $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ such that:*

- *(Non-negativity) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$,*
- *(Normalization) $\mathbb{P}(\Omega) = 1$,*
- *(Countable Additivity) Let $A_1, A_2, \ldots$ be a countably infinite sequence such that $A_i \in \mathcal{F}$ for all i and $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then*

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i).$$

(3) The probability is based on objective features. This allows us to put truth conditions on probability statements; to make probability statements true or false.

(4) *Explanation of frequencies.* There exists an intimate relation between probability and frequencies. An interpretation ideally explains the robust frequency of the occurrence of a certain outcome. For example, assume an experiment in which a fair coin is flipped 100 times, and this experiment is repeated another 100 times. Suppose that in 90 of these experiments the frequency that heads is the outcome lies between 45 and 55. An interpretation should explain why these frequencies are similar in most of the experiments; why they are robust.

(5) *Distinguishing nomic regularities.* Assume the same situation as in item (4). In 90 experiments, the frequency of heads represents the nomic, law-like nature of a (fair) coin. Now suppose that in one of the experiments the coin lands heads only 20 times. Although this is unlikely to happen, it is definitely possible, especially if you repeat the experiment many times. But this is what we would call an accident, and the frequency, 20, does not represent the nature of the coin. Therefore, an interpretation must be able to tell the difference between when such a frequency happens by accident and when it actually represents the structure of the experiment; it must be able to distinguish between nomic and accidental regularities.

4.2 Flaws of frequentism

There exist numerous different interpretations of probability, among them frequentism. However, due to lack of time, I will only demonstrate frequentism and its flaws. Since frequencies are so widely used to assign probability, demonstrating its flaws hopefully evokes enough understanding on the importance of finding a new, promising interpretation of probability.

As established in Section 4.1, probabilities are intimately connected to frequencies, and frequentism bears the most intimate relation of all: probability *is* relative frequency. Within frequentism, one can distinguish between *finite* and *hypothetical* frequentism. For example, suppose someone tosses a coin a finite number of times. A finite frequentist would then say that the probability of the coin landing heads is equal to the number of times the coin lands heads divided by the total amount of tosses; the relative frequency. A hypothetical frequentist equates probability to the relative frequency of the occurrence of an outcome in an infinite series of trials[4].

Frequentism definitely scores some points in the first two criteria; ascertainability and satisfaction of Kolmogorov's axioms. But, let us focus on its flaws:

(i) *Too absolute.* The first problem is that frequentism proposes a connection between probability and frequencies that is too absolute. It assigns probability wherever there are relative frequencies. This has several problematic consequences. Frequentism does not care whether the relative frequencies actually reflect the nomic structure of an experiment or whether they happen by accident. Therefore, they fail to distinguish between nomic and accidental regularities. In addition to its indifference towards accidental regularities, frequentism also assigns probability to events which should not even be provided with probability in the first place. It may see probability where there is none. For example, suppose that some country has elected 20 presidents since it became a democracy. And, by some accident, 8 of those were named James. The event that a president is named James is not a nomic process; someone's name is generally not connected to their intelligence and leadership capacities. However, a frequentist would assign probability $\frac{4}{10}$ to the event, implying the existence of a nomic structure where there is none.

(ii) *The reference class problem.* The main point of this problem is that any given event has more than one relative frequency. To be more precise, any given event belongs to different reference classes which yield different relative frequencies. Suppose that I want to know the probability that I am going to die before the age of 60. Well, I belong to the class of all living things, all humans, all females, all smoking females, etc. All of these reference classes carry their own (and probably distinct) relative frequency. Then, which reference class is the right one? Relative frequencies must always be relativized to a certain reference class, and there is no logically sound rule that tells us which class to choose.

(iii) *Explaining frequencies.* Frequentism does not really *explain* robust frequencies; they *identify* with them. Suppose the same situation as in the

[4]For more information on the distinction between finite and hypothetical frequentism I refer to two articles of Hàjek, (6) and (7).

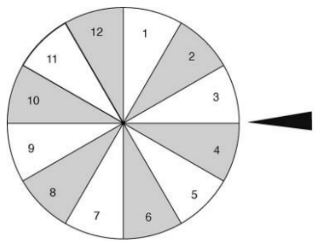

Figure 15: The wheel ((3), p.665).

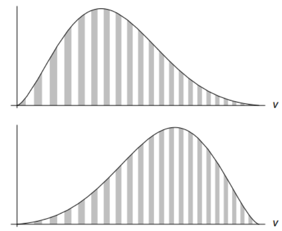

Figure 16: Two different 'reasonable' distribution functions over the initial state space ((11), p.12).

fourth criteria of Section 4.1. If I were to ask "why is the frequency of heads between 45 and 55 so often?", then the intuitive answer would probably be "because the probability is approximately equal to $\frac{1}{2}$". But for a frequentist this would translate to "because the frequency of heads is approximately equal to 50%", ending up in a vicious cycle of answers and questions. Intuitively, the probability *causes* the appearance of robust frequencies, and therefore, *explains* robust frequencies. Within frequentism, there is no causal relationship between probability and frequencies, but an identity.

For more information on frequentism and its flaws I also refer to Hàjek's articles; (6) and (7).

4.3 A promising new interpretation

Now that we are somewhat more familiar with interpretations of probability, let us take a look at the interpretation inspired by the method of arbitrary functions. In the future, the probabilities interpreted in this way will be referred to as *dynamical probabilities.* There exist multiple variations on its exact form, but they are all based on the same idea. That is, they are all dependent on the physical properties of a chance experiment and the distribution over the initial conditions. But, the most promising feature of the method of arbitrary functions is that this dependence on the initial distribution *almost* vanishes. Therefore, it *almost* solely depends on the dynamical properties, which is an objective feature. Let us take the example of the roulette wheel to illustrate this.

Assume a symmetrically balanced wheel with equally sized alternating stripes of white and gray as discussed in Section 2.6.1 (see Figure 15). Define the *initial state space* as the space of all possible initial conditions. Here, the initial state space is all positive values of the angular speed ω with which the wheel is initially spun. So, the end result, white or gray, is completely determined by ω. Therefore, putting a distribution on the initial state space allows us to compute the probability of gray (or white). Now assume that such a distribution function is ‘reasonable’, not too eccentric, just like the two distribution functions shown in Figure 16 (Strevens, 2011, p.12). Then the probability of gray, i.e. the gray area under the distribution function, is approximately equal to one over two, independently of our choice of a distribution function that is not too eccentric. Now suppose that all natural occurrences of these distribution functions are not too eccentric, then we can state that the probability of gray is approximately equal to $\frac{1}{2}$ and therefore depends only on the dynamical features of the wheel.

However, what does it mean for a distribution function to be ‘not too eccentric’? And if we know exactly what this means, how can we ensure the truth of the statement that the only way for such a distribution function to occur naturally is when it is not too eccentric? Until there is an answer to these questions we can merely say that the probability of gray is *almost* independent of the initial distribution.

You might wonder why we do not simply use Corollary 3.15 to eliminate (some of) these questions. In the limit the probability of gray is independent of any absolutely continuous initial distribution. This shows more promise because absolute continuity is a much clearer defined term than ’not too eccentric’. Moreover, absolute continuity is a much weaker condition. It is therefore much more plausible that the natural occurrence of an initial distribution is limited to absolute continuity than to functions that are not too eccentric, although the task of proving such a statement would still stand. However, in real life the initial conditions do not exceed certain values. These limiting values are hypothetical, *counterfactual*, and are therefore not relevant to us. It seems that we should stick to finite values of the initial conditions; to *actual* initial conditions.

So, the dynamical probability is *almost* determined by dynamics alone. However, as Strevens puts it, “*almost* is not worth much in the world of absolutes where metaphysics makes its home” ((11), p.15). So why should we investigate this dynamical probability anyway? Well, for multiple reasons:

(i) As we will come to see in Section 4.5, the dynamical probability satisfies Kolmogorov’s axioms.

(ii) Dynamical probability almost provides an objective interpretation of probability. But, it does require some probabilistic material to begin with, however, it requires so little of this that it is worth looking into.

(iii) Dynamical probability explains relative frequencies well.

(iv) Dynamical probability is capable of predicting relative frequencies. Consequently, it is capable of distinguishing between nomic and accidental regularities.

Thus, dynamical probability satisfies almost all the criteria listed in Section 4.1, rendering it a promising alternative interpretation of probability.

4.4 Limitations

Before we move on to the definitions of dynamical probability, it is important to note that this probability is of a certain kind and limited to certain experiments.

The kind of probability we will try to provide an interpretation of is of deterministic type. Strevens calls this kind a "deterministic probability" and defines it as "a physical probability ascribed by some scientific theory to an outcome type that is produced by processes that are, deep down, deterministic or quasi-deterministic (meaning that all relevant fundamental-level probabilities are close to zero or one)" ((11), p.1). This implies that the fact that we attach a probability to a certain outcome is due to our own limited capacities. If someone has all the knowledge relevant to the outcome, this person would know the outcome. Because there is a gap in our knowledge or physical capacities somewhere, we do not know the exact outcome, and therefore, we assign probability to it. This deterministic probability is different from the probability one encounters in quantum mechanics. Quantum mechanics obeys laws that are intrinsically chancy, meaning that even an all-knowing being, like God, is not able to predict the outcome.

As established, dynamical probability is also limited to a certain kind of experiment. In general, we will consider repeatable deterministic processes. The results of these experiments depend on the exact circumstances that obtain. Thus, all such experiments come with an *initial state space*. And the points in the initial state space uniquely fix the outcome of the experiment. This initial state space must satisfy two conditions:

(i) Small variations of these initial conditions lead to a different outcome. Or, more precisely, each not too small subregion of a point leading to a certain result contains points leading to a different result. This represents the instability of the mechanism and this is why any attempt to control the outcome is in vain and why the experiment appears random to us.

(ii) In any not too small interval of the initial state space, the proportions of initial conditions leading to a certain outcome are roughly constant. Strevens calls this condition *microconstancy*. And he calls the proportion in question the *strike ratio*. This proportion may differ for different outcomes, but it must be (approximately) constant for a certain outcome throughout the entire initial state space. Otherwise, one could manipulate the frequency with which a certain result occurs by aiming at a certain region of the initial state space which is big enough to control. In this case, the mechanism would not be able to provide a unique probability value.

Let us recall the coin toss. We saw that a certain vertical velocity, u, and a certain angular velocity, ω, determine whether the the coin lands heads or tails. Therefore, the initial state space attached to the coin toss is $\{(u, \omega) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}\}$. Suppose that one throws a coin with initial values u and ω, and these lead to heads as outcome. Then, increasing either vertical or angular

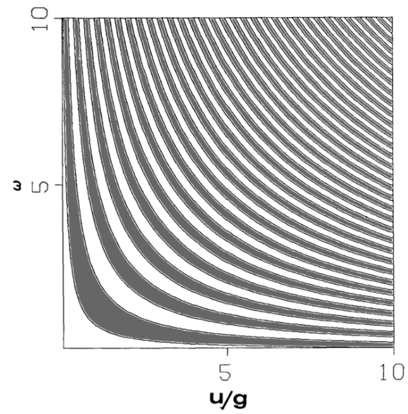

Figure 17: Initial conditions in a white area lead to heads as outcome, those in a gray area lead to tails.

velocity with a small value, δ, can lead to tails as outcome. Here, δ is so small that our hands cannot tell the difference between throwing a coin with initial values (u, ω) and initial values $(u + \delta, \omega)$ or $(u, \omega + \delta)$. We are unable to control the outcome because of our inability to control our hands in a sufficiently precise manner. Therefore, the coin toss is an experiment with an initial state space which satisfies condition (i). To see that it also satisfies condition (ii), take a look at Figure 17 again. The white and gray stripes do become smaller as u or ω increases, but all neighboring stripes are approximately the same size. So, when someone takes a not too small subregion of the initial state space, the proportion of initial values leading to heads is approximately equal to the proportion of initial values leading to tails. Therefore, the proportion of initial values leading to heads is approximately one over two in the entire initial state space, and therefore satisfies condition (ii). Thus, the coin toss is an experiment of deterministic nature which satisfies the above conditions and is therefore applicable to this interpretation of probability.

4.5 Definition

Recall that dynamical probability is any kind of probability derived through the dynamical properties of an experiment. But what the precise embodiment is of this interpretation is as yet unclear. In his article (10), Rosenthal attempts to make dynamical probability exact and calls it the *range conception*, denoted by (RC).

Definition 4.2 (RC)**.** *Let E be an experiment of the type we discussed in Section 4.4 and let A be a possible outcome of it. Let S be the initial state space attached to E, and S_A be the set of initial values leading to A as result. We assume that S and S_A are measurable subsets of the n-dimensional real vector space $\mathbb{R}^n$ (for some n). Let μ be the standard (Lebesgue-) measure. If there is a* *number p such that for each not-too-small n-dimensional (equilateral) interval I in S, we have*

$$\frac{\mu(S_A \cap I)}{\mu(I)} \approx p, \tag{68}$$

then there is an objective probability of A upon a trial of E, and its value is p.

First of all, let us check if (RC) satisfies the axioms of Kolmogorov's probability calculus (Section 4.1). The first two axioms follow directly from the nonnegativity of the Lebesgue measure and the fact that $\mathbb{R}^n \cap I = I$ for all $I \subseteq \mathbb{R}^n$. To see whether (RC) satisfies the third axiom is more difficult. Suppose that $A_1, A_2, \ldots$ is a countable sequence of pairwise disjoint sets. Define $A = \cup_{i\in\mathbb{N}} A_i$. Note that

$$S_A = \bigcup_{i\in\mathbb{N}} S_{A_i}.$$

Because a point in the initial state space *uniquely* fixes an outcome, we find that $S_{A_i} \cap S_{A_j} = \emptyset$ for all $i \neq j$. Therefore, fo all $I \subseteq \mathbb{R}^n$ it holds that

$$(S_{A_i} \cap I) \cap (S_{A_j} \cap I) = \emptyset, \quad \text{for all } i \neq j.$$

Therefore, by the countable additivity of the measure μ we obtain:

$$\begin{aligned}\frac{\mu\big(S_A \cap I\big)}{\mu(I)} &= \frac{\mu\big((\cup_{i\in\mathbb{N}} S_{A_i}) \cap I\big)}{\mu(I)} \\ &= \frac{\mu\big(\bigcup_{i\in\mathbb{N}}(S_{A_i} \cap I)\big)}{\mu(I)} \\ &= \sum_{i\in\mathbb{N}} \frac{\mu(S_{A_i} \cap I)}{\mu(I)}.\end{aligned}$$

For all $i \in \mathbb{N}$, assume that

$$\frac{\mu(S_{A_i} \cap I)}{\mu(I)} \approx p_i,$$

for each not-too-small n-dimensional (equilateral) interval I. Therefore, we can ascribe probability p_i to the event A_i. Suppose that

$$\frac{\mu\big(S_A \cap I\big)}{\mu(I)} \approx p,$$

for each not-too-small n-dimensional (equilateral) interval I and therefore we ascribe p to the probability of $\cup_{i\in\mathbb{N}} A_i$. However, this does not necessarily imply that

$$\sum_{n\in\mathbb{N}} \frac{\mu(S_{A_i} \cap I)}{\mu(I)} \approx \sum_{n\in\mathbb{N}} p_i,$$

for each not-too-small n-dimensional (equilateral) interval I. Because suppose that all p_i are on the high side, meaning that if the value of p_i is a tiny bit higher, then it would not satisfy (RC) anymore. If you take the sum of all p_i, then it is relatively seen much more on the high side, and might deviate significantly from p. Consequently, the sum of the probabilities of all A_i might not satisfy the desired conditions in order to ascribe this sum to the probability of $\cup_{i\in\mathbb{N}}A_i$. Therefore, (RC) generally does not satisfy the third axiom of Kolmogorov's probability calculus.

Rosenthal gives another similar interpretation of probability, denoted by (AF). This interpretation assumes the existence of some kind of well-behaved, not-too-eccentric, density function over the initial state space. Since an interpretation that is supplemented with some sort of notion of a density function has a tendency to be in need of another interpretation of probability, Rosenthal provides us with both (RC) and (AF) by the means of latitude.

Definition 4.3 (AF)**.** *Let E be a random experiment of the type discussed in Section 4.4 and let A be a possible outcome of it. Let S be the initial-state space attached to E, and S_A be the set of those initial states leading to A. We assume that S and S_A are measurable subsets of the n-dimensional real vector space $\mathbb{R}^n$ (for some n). If there is a number p such that for any real-valued density function δ on S that is approximately constant on (equilateral) intervals up to a certain appropriate size k, we have*

$$\int_{S_A} \delta(x)dx \approx p, \tag{69}$$

then there is an objective probability of A upon a trial of E, and its value is p.

(AF) also satisfies the first two axioms of Kolmogorov's axioms, which follows from the fact that δ is a density function. Concerning Kolmogorov's third axiom, it runs into the same issue as (RC). Due to the fact that the dynamical probability is derived through an approximation, it generally fails to satisfy countable additivity.

We see that (RC) and (AF) perform well in the first criterion listed in Section 4.1, but they do not fully satisfy the second criterion; they fail countable additivity. There are some other aspects of this interpretation that raise some questions. In Section 4.6, we will address the problem of the initial distribution. But, before we do so, the issue that becomes immediately apparent will be briefly discussed. For information on the other issues that arise I refer to Rosenthal's article, (10).

The issue that immediately sticks out is the use of vaguely defined concepts such as "not-too-small intervals" and the fact that (68) and (69) are mere approximations. Unless there is an actual realization of an experiment that is perfectly random, there will always exist a sufficiently small equilateral interval I such that all initial values contained by it lead to one and the same result. Such a perfectly random experiment can only be represented mathematically, not in real life. Therefore, we cannot remove the vague condition on the size of the intervals I. Similarly, in any experiment that does not contain the feature of perfect randomness there will always exist different intervals that satisfy the criterion of being not-too-small but which deliver slightly different values of p.

Therefore, if a certain value p satisfies (68), then there always exists an $\epsilon > 0$ such that $p + \delta$ also satisfies (68) for all $0 < \delta < \epsilon$. Therefore, it is unclear what the exact value of p is. Moreover, it is also unclear what the boundaries are on this approximation. How much is p allowed to deviate from the left hand side of (68)?

4.6 The problem of the initial distribution

Now I turn to the main problem, which is the possibility of the natural occurrence of an unusual, eccentric initial distribution. What such an eccentric distribution looks like exactly depends on the initial state space. Imagine a distribution function which appears with an extreme peak on an interval of initial values leading to the same outcome. For example, it could happen that, for some reason, someone has a tendency to flip a coin with certain initial values which happen to lead to heads. Then, the probability that heads is the outcome when that same person flips the coin would be significantly different from the one generated by (RC). The distribution of the initial values with which this particular person flips a coin is then called eccentric. We have already encountered such a distribution, namely the function in Figure 5. If such an eccentric distribution were to occur naturally, then the corresponding probability would be significantly different from the probability values prescribed by (RC) and (AF). Evidently, this poses an issue. Rosenthal attempts to resolve this in the following way. One of two things can happen. Either an eccentric distribution occurs by accident. In this case, we acknowledge that the corresponding probability is not the "true" probability and we move on. Or, such an eccentric distribution proves to be stable. In this case, we cannot simply ignore the probability it yields. Consequently, (RC) and (AF) are only applicable in certain ideal situations which renders its applicability in analysis of probability useless. To be clear, when we refer to the problem of the initial distribution, then we mean to refer to the possibility that such a specifically eccentric initial distribution arises and proves to be stable.

One option to fix this problem is to give an interpretation of the initial distribution. We could, for example, use frequencies to determine the initial distribution and use the probability that this distribution generated by frequencies yields. However, this merely pushes the task of interpreting the outcome probability back to the task of interpreting the initial distribution. We would have to start from the very beginning. So that option should be ruled out. Rosenthal and Strevens use different strategies to resolve this issue. In his attempt, Rosenthal does not change (RC) and (AF) and tries to solve this issue rationally. Strevens, on the other hand, gives a similar interpretation of probability but supplements it with something else that could potentially lead it to salvation. Strevens' approach will be discussed in Section 4.6.2.

4.6.1 Rosenthal

Let E be an experiment and let $F_1, F_2, \ldots, F_n$ be the initial conditions that determine the outcomes of E. Rosenthal argues that nature generally does not care about the initial conditions. What is meant by this is that nature normally does not favor certain particular initial conditions. If an initial distribution generates a probability that deviates from the probability that (RC) generates, then it must peak at just the right values, at very specific values that lead to a certain outcome. Furthermore, not all eccentric initial distributions generate a probability that disagrees with the one generated by (RC). So, in order to generate a deviating probability, the initial distribution must not only be eccentric but eccentric in just the right way. Moreover, small changes in an initial distribution that is eccentric in just the right way tend to reinstate probabilities according to (RC). So the conditions that a distribution must satisfy in order to generate a deviating probability are very sensitive towards disturbances. The emergence of a specific eccentric initial distribution that proves to be stable, therefore, seems all the more unlikely. Following Rosenthal's line of reasoning, the only way in which such a situation can arise is if there exists some kind of reason for its occurrence; if we overlooked other initial conditions that determine the outcome. In this case, the experiment E would not be well defined because it should involve all factors that influence the final outcome. This would imply that E does not represent the actual experiment and that the probability p that corresponds with E does not belong to an outcome of this particular experiment, but to another, albeit similar, experiment. Therefore, p is not the "true" value and the probability that corresponds with the eccentric initial distribution does not deviate from the "truth". If $F_1, F_2, ..., F_n$ were to be supplemented such that all factors that determine the result are represented, one would not encounter a stable eccentric initial distribution anymore and (RC) and (AF) would yield the true probabilities. This seems to solve the issue of the initial distribution. However, it is difficult to assess what an initial state space that involves *all* relevant initial conditions would look like. Suppose that someone flips a coin a lot of times. Usually, the ways with which the coin leaves someone's hand are taken as the initial conditions. But what if it turns out that, for this particular person, these initial conditions are eccentrically distributed? Should we then switch to a description of the entire body? Or of everything in his/her immediate environment? Apart from the fact that it is hard to define the improved set of initial conditions, it is also highly impractical.

In summary, this argument shows us that if an eccentric initial distribution occurs, then the experiment is not well-defined and the probability it generates can be dismissed. This seems to solve the problem of the initial distribution. However, the occurrence of an eccentric initial distribution requires a redefinition of the initial state space, which can be hard to assess and causes impracticality.

4.6.2 Strevens

In the previous section, it was mentioned that Rosenthal provides us with both (RC) and (AF) such that one does not have to commit to terms like 'density functions' suggesting the need for another interpretation of probability. This is what Strevens tries to do; he uses (RC) in order to interpret the outcome probabilities but avoids the interpretation of the input probability. As briefly discussed in Section 4.4, Strevens calls the probability p in (RC) the *strike ratio*. However, he supplements his interpretation of dynamical probability with something else such that the probability is no longer solely dependent on the physical structure of an experiment, but on the physical structure *plus* certain facts that guarantee the non-eccentricity of the initial distribution. The non-eccentricity of the initial distribution he calls *macroperiodicity*, which he defines as the near-flatness over small intervals of initial conditions (where what it takes to be "small" is determined by the structure of the initial state space). In this way, this interpretation assigns probability only to the outcome of microconstant (see Section 4.4) processes whose initial conditions are macroperiodically distributed. The possibility of the occurrence of an eccentric initial distribution is ruled out from the start. Therefore, this interpretation of probability avoids the main problem of Rosenthal's interpretation. However, we must investigate the consequences of such a supplementation to the interpretation. For example, what are the facts that guarantee a process' macroperiodicity? For these facts Strevens uses the frequencies that emerge in a series of *actual* initial conditions. In this way, Strevens avoids the usage of terms like 'initial distribution'. In his interpretation of probability there is no initial distribution; there are facts about initial condition frequencies which guarantee the macroperiodicity of the initial conditions of an experiment to whose outcomes he limits the assignment of probability. Strevens repeatedly stresses the fact that he does *not* give a frequentist account of the initial distribution because there is simply no space for a probability distribution over initial conditions in his interpretation.

Definition 4.4. *The outcome A of an experiment E has a deterministic probability if:*

1. *The dynamics of the experiment is microconstant w.r.t. A, and*
2. *The actual initial conditions of nearly all long series of trials on experiments of the same type as E form macroperiodically distributed sets.*

The deterministic probability, if it exists, is stipulated to have a value equal to the strike ratio for A.

Call the deterministic probability defined above *microconstant probability*. The conditions under which microconstant probability is stipulated contain some vague terms, such as "nearly", "all long trials", and "of the same type". For a clarification of these terms, I refer to the article (11). Another matter needs clarification, because what does it mean for actual initial conditions to be macroperiodically distributed? Divide the initial state space up in small subregions, where small is determined by the notion of small in the definition of microconstancy. For all these subregions determine the frequency with which the actual initial conditions lie within the corresponding subregion. Summarize the results by a plot of points and connect the dots. The function that appears is the "density" function of the initial distribution. If this function is approximately macroperiodic, then the initial distribution is macroperiodic.

Although Strevens' interpretation seems to solve the problem of the initial distribution, it runs into some other issues. Going into these would mean that we lose focus, because they do not concern the problem of the initial distribution. I will only say the following. As established in Section 4.3, one of the promising features of dynamical probability is that it is able to distinguish between nomic and accidental regularities, whereas frequentism fails to do so. Strevens even calls this the *fundamental* flaw of frequentism. In his article, he states to believe that "many of the standard objections to frequentism have this dissatisfaction at their core" ((11), p.5). But, and here comes the crux, a supplementation of this deterministic probability with a frequentist component seems to get rid of this promising feature, and "renders this dynamical approach no better than straight frequentism" ((11), p.3).

Initially, Strevens seems to solve the problem of the initial distribution by ruling out the occurrence of an eccentric initial distribution right from the very start; in the definition of dynamical probability. However, the consequences of this move seem to be too severe for it to be considered as a true solution.

5 Conclusion

The method of arbitrary functions assigns dynamical probability to certain outcomes of certain experiments. We have found that dynamical probability depends on both the physical properties of an experiment and the distribution of its initial conditions. In order to consider this kind of probability as an *alternative* interpretation of probability, the dependence on the initial distribution must somehow be eliminated. Otherwise, one must interpret this initial distribution which would bring us right back to the task of interpreting probability.

The method of arbitrary functions provides one way of eliminating this dependence. The idea behind the method of arbitrary functions is that as some kind of limit is taken, the dynamical probability becomes independent of the initial distribution, as long as it belongs to a certain class of functions. In Chapter 2, we discussed the method of arbitrary functions applied to the coin toss. Theorem 2.2 showed us that the probability of heads approaches one-half as the support of the density function over the initial vertical and angular velocity shifts to infinity, as long as this density function is continuous. In Chapter 3, we generalized the theory behind the method of arbitrary functions such that it applies to all kinds of deterministic experiments, and not just to the coin toss. However, the method of arbitrary functions only eliminates this dependency in the limit. We have observed that actual initial conditions do not take on infinity values. Therefore, the method has no applicability in real life.

It has, however, inspired philosophers to think of probability as the physical properties of an experiment; it has inspired an alternative interpretation of probability. It turns out that if dynamical probability solely depends on the physical properties of an experiment, then it contains some promising features. Many are motivated to investigate this alternative interpretation because there exists a lot of ambiguity around what probability is exactly, and all suggestions that exist at this point meet significant resistance.

But, as I said, dynamical probability is only able to claim most of its promising features if it is solely dependent on an experiment's physical properties. Therefore, we must first think about how we can reach this state. So, we went to find a solution to the *problem of the initial distribution*. We have discussed two possible solutions. Firstly, it can be argued that the occurrence of an eccentric initial distribution only occurs if the experiment is not well-defined. Therefore, we can dismiss the probability it generates. Second, the possibility of the occurrence of an eccentric initial distribution can be ruled out from the start; we limit the assignment of dynamical probability to experiments whose initial distribution is somehow ensured to be non-eccentric. The consequences of this move, however, seem to remove one of the main promising features of dynamical probability; distinguishing between nomic and accidental regularities. One can therefore wonder whether this solution is worth the trouble.

In conclusion, we have found no proof that eccentric initial distributions do not occur. Therefore, whether the problem of the initial distribution is solved remains ambiguous.

References

[1] Abrams, M. (2012). *Mechanistic probability.* Synthese, 187, p.343-375.

[2] Billingsley, P. (1995). *Probability and measure* (3rd ed.). John Wiley & Sons.

[3] de Canson, Chloé. (2022). *Objectivity and the method of arbitrary functions.* The British Journal for the Philosophy of Science, 73, p.663-684.

[4] Engel, Eduardo M.R.A. (1992). *A road to randomness in physical systems.* Lecture notes in statistics, 71. Springer-Verlag, Berlin.

[5] Hopf, Eberhard. (1934). *On causality, statistics and probability.* Journal of mathematics and physics, 13, p.51-102.

[6] Hàjek, Alan. (1997). *Fifteen arguments against finite frequentism.* Erkenntnis, 45, p.209-227.

[7] Hàjek, Alan. (2009). *Fifteen arguments against hypothetical frequentism.* Erkenntnis, 70, p.211-235.

[8] Hájek, Alan. (2023). *Interpretations of Probability.* The Stanford Encyclopedia of Philosophy (Winter 2023 Edition). URL = https://plato.stanford.edu/archives/win2023/entries/probability-interpret/.

[9] Keller, Joseph B. (1986). *The probability of heads.* American Mathematical Monthly, 93, p.191-197.

[10] Rosenthal, Jacob. (2012). *Probabilities as ratios of ranges in initial-state spaces.* Journal of Logic, Language and Information, 21, p.217-236.

[11] Strevens, Michael. (2011). Probability out of determinism. *Probabilities in physics* (p.339-364). Oxford University Press.

[12] von Plato, Jan. (1983). *The method of arbitrary functions.* British Journal for the Philosophy of Science, 34, p.37–47.