Ludwig-Maximillian-University

Master Thesis

# Randomness, a point of view in dynamical systems with applications in statistical physics and number theory

*Author:*
Han Yu

*Supervisor:*
Prof. Peter Pickl

*A thesis submitted in fulfilment of the requirements for the degree of Master of TMP*

*in the*

Master Program
TMP

November 2014

# Declaration of Authorship

I, Han Yu, declare that this thesis titled, 'Randomness, a point of view in dynamical systems with applications in statistical physics and number theory ' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a master degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

LUDWIG-MAXIMILLIAN-UNIVERSITY

# *Abstract*

Faculty Name
TMP

Master of TMP

**Randomness, a point of view in dynamical systems with applications in statistical physics and number theory**

by Han Yu

Randomness is a long time unclear concept in science. Although in probability theory we have Kolmogorov's axioms as a rigorous framework, there are deep conceptional problems in understanding randomness. Think, for example, on a deterministic dynamical system. Although everything is determined by the initial condition, such systems can be used to model random experiments. In this thesis a dynamical point of view of probability is discussed. We will also discuss interesting applications of dynamical systems in number theory, for example Szemeredi's theorem. From the aspect of algorithmic information theory the concept of randomness of a sequence of codes is highly correlated to its incompressibility which deals with some algorithmic features. This will also be discussed in this thesis

# *Acknowledgements*

1.The existence of the universe ensure us to understand and argue.

2.The existence of TMP helps me to get the chance to do 1).

3.The existence of my supervisor Prof. Peter Pickl helps me in this certain topic on randomness.

4.The existence of my friends makes this thesis finally finished.

# Contents

# List of Figures

# List of Tables

## 0.1 Randomness discussed in this thesis

The main topic of this thesis is about randomness, however, this concept is not clear even today. There are several different interpretations in different research fields and regardless of the ambiguity of the randomness used in most of the research fields, the results we have are usually quite convincing and can be tested in real life. Let's take Gibbs formalism in statistical physics as an example. Without doubts, it gives us very nice results about the properties of some complex systems. The range of application of this formalism goes from simple ideal gas system to fluid dynamics, and even further to financial systems. The reason why we could use such formalism in so many different situations is not yet clear, and sadly, I'm not able to give an answer in this thesis. But let's start from the basic foundation of the Gibbs formalism. The main point is not necessary the special form of Gibbs measure on some certain sets (usually called phase spaces), but the fact that we do introduce some measure with some certain relation with the dynamics of the phase space (measure preserving). And then identify the quantities we can observe or measure by the functions on phase space and treat them naturally as random variables in the sense of probability theory. We can calculate the mean value, second order correlations, higher order correlations (from this point the full procedure of "diagram" expansions comes into being) , we can study the limiting properties with respect to some parameters and get the thermodynamic limit. The formalism is mathematical complete, however, there is an open question: Why do we have to introduce the measure?

In view of probability theory we can treat the measure as probability measure whenever it is finite and non-negative. If we want, we can even introduce the generalized probability as real-valued or even vector-valued measures. Then we adapt our mind on measures as probabilities and through out the whole theory of statistical physics we always interpret the measure as probability measure and functions as random variables. That is nice as it is, but the question in the last paragraph is still open: Why do we have to introduce the probability?

To make the point clear, let's notice that probability theory as a mathematical theory is nothing more than finite measure analysis. In the logical point of view we can one-to-one translate probability theory to finite measure analysis by changing the language we use (probability to measure, random variables to measurable functions and so on).

Now let's formulate the question in the first paragraph in a clear way:

Consider a dynamical system (definitions will be given later, but for now we can think it as Hamiltonian systems). We have two different descriptions, namely, on one hand the dynamics (for Hamiltonian systems the dynamics are the Hamiltonian flows) fully describing the system, and on the other hand some probabilistic descriptions. These two descriptions of the same systems have to be compatible. Then it is not avoidable that we have to deal with two questions:
1. What is the meaning of probability for dynamical systems which is fully characterized by dynamics?
2. After we have understood the meaning of probability, are the two descriptions really compatible?

If these two questions are solved we will have a well founded statistical physics without any problems in interpretations.

## 0.2 Measurable dynamical systems

Instead of immediately trying to answer the questions in the end of last section, let's first check what kind of object we are dealing with in those questions. First we have a dynamical system, which will be a set $\Omega$ called phase space together with a dynamics which can be defined as an indexed self mapping $T^I$ with semi-group property on the index. Hamiltonian systems are such kind of object. There we have a phase space in the usual sense and the Hamiltonian flow as the dynamics (with a real time as one parameter group). Then we have the probabilistic description which requires us to make the phase space $\Omega$ finite-measurable. The object with such properties is called finite measurable dynamical systems. We see that in order to deal with the questions in the end of last section we have a minimum requirement of the object we are dealing with – the measurable dynamical system (usually whether the measure is finite or not is not a problem, but what really matters is the structure of the measurable sets).

Now, if we want to interpret probability in frequency sense (as people usually do in physics), we should study the conditions under which probability is compatible with dynamics. That is a difficult part and, up to now, the best general results are the concept of ergodicity. Ergodicity is not a necessary condition and it is practically very difficult to be checked. But it is a standalone mathematical topic with a strong power on its own.

The first two chapters will be about the basic concepts of measurable dynamical systems and the ergodic theory together with some applications in number theory. The third

chapter will deal with the old issues on Boltzmann's statistical physics and the objections of Loschmidt and Zermelo. In the forth chapter we will discuss some certain ergodic dynamical systems and illustrate the connection between probability and its frequency interpretation.

## 0.3 Unfinished ideas and future projects

There are some unfinished works in the thesis, which are:
1. I wanted to set up a duality between the probabilistic formalism and the deterministic formalism: for every theory with probability involved we can set up a dynamical system to make the probabilistic theory a deterministic one. So far I can only find some results on the stochastic processes which are Markov which are nothing new to us (just think of the simulations of stochastic processes).
2. In chapter 4 I introduce the probabilistic way of solving 1d diffusion equations. This can be generalized to Harmonic functions. To this point I wanted to generalize the results further. We can represent the some kinds of differential equations with stochastic processes, it can be treated as the path-integral method on solving PDE's.
3. In chapter 2, I introduce the Szemeredi's theorem and a generalization which is not very easy to work with. I wanted to derive another equivalent condition which is easier to check.

# Chapter 1

# On the Gibbs formalism of general dynamical systems

Here most of the ideas and formalisms can be found in [1, 2] together with some concrete examples.
As we are talking about statistical physics, then in most cases we will discuss the Hamiltonian systems, while not, we still use some similar formulations, the Ising model for example is not Hamiltonian system, it is not even a dynamical system, all the things we have is a certain measure, so if we want to talk about something dynamical it is very intuitive that we take at least a dynamical system, in this chapter I will apply the Gibbs formalism on general dynamical systems and discuss some interesting properties.

## 1.1 The basic mathematical settings

**Definition 1.1.** By saying a general dynamical system $(\Omega, T^I, \mathcal{F})$ means that we have a measurable space $\Omega$ with $\sigma$-algebra $\mathcal{F}$ and a family of self-mapping $T^I$ indexed by the set $I$ which is assumed to be a semi-group discrete or continuum(means the cardinality of the set I is either at most countable or the same as $\mathbb{R}$)and the family of self-mapping $T^I$ form a one parameter semi-group.

Usually, $\Omega$ is called the phase space and $T$ is called the dynamics. Classical mechanics is a perfect example of such object, where the phase space $\Omega$ is been understood in a usual way and the dynamics $T$ in this case is the Hamiltonian flow.

Here no further information of the measure on $\Omega$ and the connection properties between the self-mapping family $T^I$ is required, just to keep the notion general and when needed

i will clarify those properties.

**Definition 1.2.** By saying that the family of self-mappings $T^I$ is measurable means that:
$\forall i \in I$ and $\forall A \in \mathcal{F}$ we have $(T^i)^{-1}(A) \in \mathcal{F}$.
By saying that the measurable family of self-mappings $T^I$ is non-singular with respect to a given measure $\mu$ on $\Omega$ means that:
$\mu((T^i)^{-1}(A)) = 0$ whenever $\mu(A) = 0$ provided that $A \in \mathcal{F}$.
By saying that the measurable family of self-mappings $T^I$ is invariant with respect to a given measure $\mu$ on $\Omega$ means that:
$\mu((T^i)^{-1}(A)) = \mu(A)$ provided that $A \in \mathcal{F}$.

It is clear from the definition that the invariance implies the non-singularity.And the non-singularity can be seen as follows if we introduce a result from measure theory:

**Theorem 1.3.** *If we define a new measure $\lambda$ by:$\lambda(A) = \mu((T^i)^{-1}(A))$ where $T^I$ is non-singular then measure $\lambda$ is absolutely continuous with respect to the measure $\mu$ in the sense that $\forall \epsilon \gneq 0\ \exists \delta \gneq 0$ such that$\lambda(A) \lneq \epsilon$ whenever $\mu(A) \lneq \delta$*

The proof of the theory with some generalization can be found in [3] Theorem III.6.15

## 1.2 The Gibbs formalism

The Gibbs formalism(or Gibbs theory,Gibbs method,Gibbs approach) on statistical physics is mainly about the ensembles,namely,when taking Hamiltonian system for example,we give the phase space often denoted by $\Gamma$ a measure which is absolutely continuous with respect to the Lebesgue measure which is invariant under the Hamiltonian flow by the theorem of Liouville and moreover the measure(canonical, micro-canonical) are all invariant under the Hamiltonian flow.If we do not ask sharp questions on the physical meaning of this formalism, it is a well-defined mathematical structure which this framework it is possible to formulate the thermodynamic limit in a very precise manner.
But however,Gibbs formalism faces a hard time in finding the description of non-equilibrium,a question often appear is that what it means by saying the system approach to equilibrium,this question is something related to the dynamics of the system,and there are much to say about that both philosophically and mathematically.
Here i follow a naive idea that if we identify the state of a system by "distribution"

namely,some general ensemble on the phase space,then it is possible to understand and study the change of the "distribution" under the dynamics,and it is also possible to formulate the "approach to equilibrium" in this sense.Philosophical discussions will be given later.

## 1.3 The induced dynamics of the distribution

As stated above,the "distribution" is in some sense identified with the state of the system,this identification is debatable of course but nevertheless in some books for example [4] both on mathematical aspect and physical aspect it is assumed the by saying the state it means some certain distribution.

**Definition 1.4.** By saying a distribution of the dynamical system $(\Omega, T^I, \mathcal{F})$ it means a non-negative finite measure $\mu$ on the space $\Omega$.We can further normalize the measure to a probability measure.

Then from a theorem from Radon-Nikodym we can identify the set of all the normalized measure absolutely continuous with respect to $\mu$ with the set of all $L^1$ functions with $L^1$ norm 1.
Again from that theorem,if we have a non-singular family of self-mappings $T^I$ for each $i \in I$ we can define a absolutely continuous measure with respect to $\mu$ and in addition it is normalized since $(T^i)^{-1}(\Omega) = \Omega$ Which means we can find a $L^1$ function $f_i$ correspond to the new measure.
From this we can see that the dynamics $T^I$ induces a dynamics on the $L^1$ space.However if we want to introduce the dynamics related dynamics of the $L^1$ space there is another choice namely we can define $T^i * (f)(\omega) = f(T^i(\omega))$ later we will see the relation between these two ways of defining change of $L^1$ functions.

**Definition 1.5.** Let $(\Omega, T^I, \mathcal{F})$ be a dynamical system with measure $\mu$ on $\Omega$,then for every $L^1$ function $f$ with non-negative value we can define $f_i$ for each $i \in I$ by Radon-Nikodym theorem.The new function $f_i$ is still in $L^1$ and $\int_A f_i d\mu = \int_{(T^i)^{-1}A} f d\mu$ for every $A \in \mathcal{F}$ it is unique up to the value on a zero measure set.Denote this transformation of $L^1$ functions induced by the family of self-mappings $T^I$ by $T^{I*}$.
With the same settings,the transformation $T^i * (f)(\omega) = f(T^i(\omega))$ is denoted by ${T^I}_*$

Notice that the definition using Radon-Nikodym is just the same as the conditional expectation in probability theory.

Here are some properties of $T^{I*}$:

**Lemma 1.6.** *For any $i \in I$ the $T^{i*}$ is linear,positive and $L^1$-norm preserving operator,and for different $i \in I$ it is a one parameter semi-group,which is called the Frobenius-Perron operator associated with the dynamics $T^I$.Simply denoted by $P^I$.*
*The ${T^I}_*$ is also linear and positive and form a one-parameter semi-group act on any functions in particular $L^\infty$ and it is often called the Koopman operator.*
*The Koopman operator and the Frobenius-Perron operator as seen as operators act on the mutual-dual spaces,are adjoint to each other.*

The proof can be found in [1] together with some concrete examples of the explicit form of the operator for some certain dynamical systems.

The intuitive idea behind this is that the probability of some set in current time comes from the probability of the set in some previous time.

## 1.4 Ergodicity,mixing,exactness and the evolution of the distributions

Before the discussion of this section,i will have to say that this is just some mathematical formulation which is also interesting in some other mathematical sense but it is not necessarily related to any physics, there a lot of disagreement on this approach to non-equilibrium physics which i will state later.
Since the proofs in this section are not trivial with some technical method which are not related to the topic under discussion and moreover they can be found in [1], [2], [5] and other books on the topic general dynamical systems,it makes no sense to copy them here since i want to focus on the overall structure of this formalism.

Nowadays there are different definitions about the ergodicity,mixing, exactness and so on,those definitions are all equivalent otherwise they will not at the same time be the definitions of a same concept,detail discussions are hold in [5], [6], [7] as well as [1].

By simplicity here for the family of self-mappings i consider the discrete case,for contiuous case or even more general semi-group with a lattice structure the formulation are

similar.

**Definition 1.7.** Let $(\Omega, T, \mathcal{F})$ be a dynamical system with a measure $\mu$ on $\Omega$ and the family of self-mappings is non-singular with respect to $\mu$.
By saying that the system is ergodic it means that whenever we have a invariant set with respect to $T$ we have either $\mu(A) = 0$ or $\mu(A) = 1$.
By saying that the system is mixing it means that $\mu$ is invariant and moreover whenever we have two measurable sets $A$ and $B$ we have $\lim_{n\to\infty} \mu(A \cap T^{-n}(B)) = \mu(A)\mu(B)$.
By saying that the system is exact it means that $\mu$ is invariant and moreover for any measurable set A we have $\lim_{n\to\infty} \mu(T^n(A)) = 1$

Notice that it is possible to see that ergodicity is strictly weaker than mixing which is againg strictly weaker than exactness.

Since in all the definitions above the non-singularity is assumed,which means it is possible in all those cases to define the Frobenius-Perron operator,and indeed all the properties of the dynamical systems listed above can be translated to the properties of the operators.To be precise,the existence of the fix points,the uniqueness of the fix point,the limiting properties towards fix points.

**Theorem 1.8.** *If the system is ergodic then the only fix point of the Frobenius-Perron operator is almost everywhere constant,in the case of probability measure the constant is 1,and the converse is also true.*
*If the system is mixing then the Frobenius-Perron operator has the property,for all distribution functions $f$,we have $\lim_{n\to\infty} T^{n*} f = 1$ in the weak sense and the converse is also true.*
*If the system is mixing then the Frobenius-Perron operator has the property,for all distribution functions $f$,we have $\lim_{n\to\infty} T^{n*} f = 1$ in the strong sense and the converse is also true.*

These three equivalence can be seen also as:

**Theorem 1.9.** *If the system is ergodic then the only invariant measure which is continuous with respect to $\mu$ call it $\lambda$ is such that*
$\lambda(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{i=n-1} \mu(T^{-n}(A))$
*If the system is mixing then the measure $\mu$ is the only invariant measure and if we*

*have any initial measure continuous with respect to it, then the dynamics induced by Frobenius-Perron operator will weakly converge to $\mu$*
*If the system is mixing then the measure $\mu$ is the only invariant measure and if we have any initial measure continuous with respect to it, then the dynamics induced by Frobenius-Perron operator will strongly converge to $\mu$*

The construction of the Frobenius-Perron operator with the help of Radon-Nikodym theorem seems to be very abstract but however if we take a deep insight into the construction we can see that if we get a new distribution function at some time n from the old distribution function at time 0,then the probability measure induced by the distribution is such that at time n the new probability measure of a set $A$ is the old probability measure of set $T^{-n}(A)$,in some physics sense,for example in the construction of Boltzmann equation or in general Liouville equation in statistical physics the idea of the change of the distribution in phase space is exactly the construction of Frobenius-Perron operator in continuous time case.
However nice and simple the mathematical structure this formalism is,it has no good use inside physics,so far the only Hamiltonian system contains arbitrary number of particles which is proven to be ergodic is the ideal gas system contained in some regular box called billiard see [8].Then what we want to ask is that what if we are dealing with some other systems,for example gas with interactions,or we want to understand the property of the system when the number of particles is large.From the basic setting of this formalism we can see that properties such as ergodicity,mixing or even exactness is what we call structural unstable,namely,it depends highly on the form of dynamics,in the discussion here is that the form of interaction matters a lot.And this formalism has a starting point of a given dynamical system,namely,different particle number will give different systems,in this sense we can not study any properties dealing with the tendency when the number of particle is large.
But nevertheless,this formalism is a precise mathematical object,even facing difficulties in describing the real physics systems,it has some mathematical interest itself.

Here is just the beginning of the discussion of dynamical systems, we do not need to treat them as real physical systems, for example ergodicity is quite a mathematical clear but physical useless property, however all the formalism above took the ideas from physical systems, but the real physical systems are so complicated that this kind of approach has a limit of reachability in the sense we can not really check any dynamical properties introduced above easily for real systems, but nevertheless the formalism itself is very rich in mathematics, and in the next chapter we will see a interesting application.

# Chapter 2

# Some words on general dynamical systems

As appeared in the discussion of although non-physical Gibbs formalism of general dynamical systems,there are several interesting aspect of the properties such as ergodicity and mixing.

The concept of ergodicity was first introduced by Boltzmann himself in order to fight against some disagreement of his statistical mechanics,on of which is the Poincare recurrence "paradox" which implies that for a general dynamical system with finite measure,then arbitrarily chose a measurable set,then almost all the points in this set will return this set even infinitely many times,for the precise discussion of this Poincare recurrence lemma see [6],Boltzmann replied with the idea that,it is true that our system will return back to its origin or infinitely near to origin,but the time we have to wait is very long,much longer than the time scale we are interested in the system.

An interesting mathematical model for this argument is the so-called Kac ring model see [9],which gives an explicit discrete dynamical system,and with a argument that for most of the initial conditions,the system will approach to equilibrium and stay for long time,but it will return to non-equilibrium in finite time which is much larger than compare to the time tend to equilibrium.

Borrow this into realistic system,we can say(guess) for example a drop of ink in the water,when we consider the drop of ink in the water as a Newtonian system with different color hard balls(which fulfill the condition of Poincare recurrence) then for "most of"(in the sense of measure,usually use the Maxwellian distribution on the velocity space) the initial conditions the drop of ink will spread over the water and reach a equilibrium when the ink is near uniformly distributed and stay for a very long time such that the time when we can observe the system form back to a drop of ink will be too long such that no one can even observe it(for example longer than the age of universe).However,notice

that the system of drop of ink in water when considered as Newtonian system of hard balls is a well defined system which in principle we can calculate everything we want,but such work is very difficult not to mentions that it will give any convincing result,because then real system is not necessary Newtonian system with hard balls,there are complex interactions or even quantum effects.
But this can not prevent us from talking about ergodicity as an interesting mathematical concept.Boltzmann's idea is that the system will go through most of the possible states before it returns,and since the phase space will be very large when we have large number of particles the recurrence time is considerably long.
Notice that the definition of ergodicity introduced in the previous chapter is different than the Boltzmann's original idea,but as we will see,under some conditions they will be compatible.

**Theorem 2.1.** *Let $(\Omega, T, \mathcal{F})$ be a general dynamical system,in addition $\Omega$ has a topology structure,and the measure $\mu$ on $\Omega$ is regular with respect to the topology,in the sense that $\mathcal{F}$ contains the Borel sets induced by the topology and for any measurable set $A \in \mathcal{F}$ we have $\mu(A)$ be the supremum of the measure of the compact subsets of $A$ and the infimum of the measure of open sets containing $A$.And non-empty open sets has positive measure. Then if the system is ergodic and continuous and open and we assume the $\mu$ to be the invariant measure(we can do that because the regularity is preserved if we change the measure to another one continuous with respect to it),then the only continuous real valued functions invariant under the dynamics induced by the Koopman operator are constant function.And all the trajectories will be dense or no where dense.And the trajectory of any open set is dense.*

*Proof.* The first statement is easy to see since the ergodicity implies that then invariant functions are constant almost everywhere,with continuity the zero measure set on which the function is not equal to that constant is open thus it is empty.
For the second statement,notice that any trajectory is an invariant set,since the dynamics is continuous the closure of the trajectory is also invariant,so is the complement of the closure of the trajectory denoted by $C$ which is open.So there are two cases namely $\mu(C) = 1$ or $\mu(C) = 0$,in the first case we have $C'$ is no where dense,in the second we have $C = \phi$ so $C' = \Omega$,which are precisely the two cases in the statement of the theorem.
For the last one,notice that the trajectory of any open set will be a union of open sets thus open,it is invariant so it has either measure 1 or 0,it can not has measure 0 since it is a non empty open set,so it has the full measure,which means it will be dense. □

## 2.1 ergodicity,some interesting results

There is a very important results about the ergodic systems connected with a theorem from Birkhoff.For proofs see [3] Chap. 8

**Theorem 2.2.** *For a measure preserving dynamical system* $(\Omega, T, \mathcal{F})$ *with measure* $\mu$ *and for any real valued* $L^1$ *function* $f$*,we have the limit*
$f^*(\omega) = \lim_{n\to n} \frac{1}{n} \sum_{i=0}^{i=n-1} f(T^i(\omega))$.
*exists for almost every* $\omega \in \Omega$ *and in addition* $f^* \in L^1$
*More over if the system is ergodic with measure* $\mu$ *assume to be the ergodic measure(the invariant measure of ergodic system) then we have in addition* $f^*$ *is almost every constant and this constant is equal to* $\int_\Omega f d\mu$

Notice that once we have a topological dynamical system and we are interested in continuous functions then the almost everywhere can be replaced by everywhere.For an mathematical interesting application of this theorem see [10],the Furstenberg's ergodic approach to Szemeredi's theorem which has then an important application to Green-Tao's theorem on the existence of the arbitrary long arithmetic progressions of prime numbers see [11].
The application in number theory seems too mathematical,but however non-physically relevant it seem to be the idea of ergodicity and all those formalism with dynamical-theoretic description have some surprisingly non-trivial applications.One of which i will talk about later on the stochastic modelling.

## 2.2 More on recurrence

As already mentioned before, a recurrence theorem by Poincare was used by Zermelo to object Boltzmann's H-theorem. The Poincare recurrence theorem can be stated as:

**Theorem 2.3.** *For a measure preserving system* $(\Omega, T, \mu)$ *with finite measure* $\mu$ *on* $\Omega$*, let* $A \subset \Omega$ *be a measurable set, then for almost every point* $\omega \in A$ *there exist a number* $n(\omega)$ *such that* $T^{n(\omega)} \in A$*, where "almost every point in* $A$*" means that the set of points which do not have the property above is measure* 0 *by* $\mu$*.*

Here the dynamical system is a time discrete system, the theorem also holds for time continuous system by observing that for any time continuous system $(\Omega, T^\tau, \mu)$ we can

get a new time discrete version $(\Omega, T^1, \mu)$ and the recurrence property for a specific discrete time implies the recurrence in general.
The proofs can be found in [5].
Now let's apply this theorem to Hamiltonian system of particles moving in a finite space with finite measure on the velocity space (This condition is satisfied for micro-canonical measure for systems with bounded energy surface), let's start the system at time 0 with a certain initial condition, denote this initial condition by $\omega_0$ then we can find a ball on the corresponding energy surface denoted by $B_{\omega_0}$ with arbitrary small radius, then what we know from the theorem is that there exists a point in that ball $\omega \in B_{\omega_0}$ which will return back to $B_{\omega_0}$ after some time $t(\omega)$, since we are now discussing a continuous system which means that there is a ball around $\omega$ denoted by $B_\omega$ with the property that $H^t(\omega)(B_\omega) \subset B_{\omega_0}$ where $H^t$ is the Hamiltonian flow. So whenever there is a recurrence point $\omega$ there is a recurrence open neighbourhood around that point with at most the same recurrence time $t(\omega)$, we know also that the set of recurrent points has the same measure as $B_{\omega_0}$, then by the regularity of the measure with respect to the topology (restricted to the energy surface), we can say that all the points in the ball $B_{\omega_0}$ are recurrent and in particular $\omega_0$ itself. Since in the beginning we can choose the ball as small as we want, which means that our system will return back to its origin as close as we want.
Now we know that the system will return near to its origin, and Boltzmann's H-theorem states that for a certain function the value of the function induced by the Hamiltonian flow is never decreasing, the function constructed is also continuous except on some measure 0 set, then we can see that the function can not be increasing as well. So the argument here for entropy (which is related to the certain function constructed by Boltzmann) increasing is not valid. Or we get the conclusion regardless of the initial conditions our the H-function will be always constant, but Boltzmann also stated that for there exist the situation where we have strictly increasing H-function, then we have a contradiction.
By logic, we know whenever we have contradictory statement under some logic structure, then they can not both be true, otherwise our logic structure is not self consistent and can be used to derived everything which is absurd. If we look back to Boltzmann's argument of the H-theory we can find the problem. The H-theorem is based on Boltzmann's equation, which is the change of distribution induced by Hamiltonian flow UNDER SOME APPROXIMATIONS (to make this point i will open another chapter to discuss that), means that the Boltzmann's equation is an approximation of the Hamiltonian system, which is not exact up to arbitrary order of preciseness, and if we describe the system using Boltzmann equation then after some time it will not be a proper description. However the Zermelo's objection is based on the exact Hamiltonian flow, and it gives the recurrence property, so the objection of Zermelo does not really contradict

with Boltzmann's approximated H-theorem. (Similarly we can also use this argument on the Loschmidt's objection)

So we see that how mathematics turned physics down, although in this case there is actually no contradictory. Recurrence property is a interesting topic of dynamical systems but maybe not in the physics sense, in fact we have some worse behaver of Hamiltonian system, first we can fix some small number $\epsilon$ and introduce the time discrete system $(\Omega, T^\epsilon, \mu)$, then we can see that if we observe the system in a not very proper time there will be a very strange behaver in the following sense:

**Lemma 2.4.** *For the system $(\Omega, T^\epsilon, \mu)$ stated before can we start the system at an arbitrary point $\omega$, then let $B_\omega$ a ball with small fixed radius, then for every nature number $k$ there exist nature numbers $M(k), d(k)$ such that $T^{M\epsilon}(\omega), T^{(M+d)\epsilon}(\omega), T^{(M+2d)\epsilon}(\omega)....T^{(M+(k-1)d)\epsilon}(\omega)$ are all inside $B_\omega$*

This lemma follows from a more general theorem by Furstenberg which will be stated later, is called the multiple recurrence property.
From this we can see that if we observe the system in some successive time points with equal delay then the system will seem to be frozen, but it is not really physically interesting since the waiting time for this to happen maybe very long.
The Furstenberg's theorem can be seen as some extension of Poincare's recurrence theorem.

**Theorem 2.5.** *For a measure preserving system $(\Omega, T, \mu)$ with finite measure $\mu$ on $\Omega$, let $A \subset \Omega$ be a measurable set with positive measure, then for any natural number $k$:*
$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{i=n} \mu(A \cap T^{-i}(A) \cap T^{-2i}(A) \cap T^{-i}(A)... \cap T^{-ki}(A)) > 0$

Notice that the positivity of the liminf will guarantee the existence of a positive term of form $\mu(A \cap T^{-i}(A) \cap T^{-2i}(A) \cap T^{-i}(A)... \cap T^{-ki}(A))$ for some nature number $i$, the apply this result to Hamiltonian system we can get the result of the lemma before.

The whole proof of this theorem will be not suitable to be included here, for the original proof see [10], for the proof with fully motivated examples see [12].

## 2.3 Furstenberg's theorem and the Szemeredi's theorem

From Furstenberg's theorem we can derive an interesting number theory theorem which was conjectured by Erdös and Turán, which stated that a set of number with some

certain property contains arbitrary long arithmetic progressions, actually the conjecture was state as:
1.(unsolved)For $A \subset \mathbb{N}$ if $\sum_{a\in A} \frac{1}{a}$ diverge then $A$ contain arbitrarily long arithmetic progressions.
2.(solved) For $A \subset \mathbb{N}$ with upper Banach density it contains arbitrarily long arithmetic progressions

**Definition 2.6.** For $A \subset \mathbb{N}$ with upper Banach density it means that,
$\limsup_{n-m\to\infty} \frac{|A\cap[m...n]|}{n-m} > 0$

The first one is still unsoved and the second one was proved by Szemeredi and it is now called Szemeredi's theorem, it is actually possible to prove Szemeredi's theorem by Furstenberg's theorem.( here i followed Einsiedler)

*Proof.* Consider the space of all one side infinite 0,1 sequence $\Omega$ with dynamical shift $T$, then it is possible to define a sequence with a property $\omega_i = \chi_{i\in A}$ where $A$ is a subset of natural numbers with upper Banach density. Call this sequence $\omega$.
It is possible to introduce the topology on $\Omega$ as the product topology of the discrete topology of $\{0, 1\}$, which will make $\Omega$ compact and metrizable.
Now take the sequence $\omega$ which is a point in $\Omega$ and take the closure of the trajectory $Tra(\omega) = cl\{T^n(\omega), n \in (N^+)\}$.
Now the closed subset of $Tra(\omega) \subset \Omega$ is also compact and metrizable, so if we introduce the Borel algebra on $Tra(\omega)$ which can be seen as the Borel algebra restricted on $Tra(\omega)$, then the space of probability measure will be weak*-compact since it is actually a closed subset of the unit ball of the dual space of $C(\Omega)$.
Now we can define a open subset of $Tra(\omega)$ as:
$U = \{\omega' \in Tra(\omega)|\omega'_0 = 1\}$.
Let $\delta_m$ denote the singular (Dirac) measure of point $T^m(\omega)$, the it is possible to see that: the measures of the form:
$\frac{1}{b-a}\sum_{m=a}^{m=b} \delta_m$ for $a, b$ positive natural numbers will contain limit points by the weak*-compactness of the space of probability measures and we have that there will exist a subsequence of $\frac{1}{b-a}\sum_{m=a}^{m=b} \delta_m$ with
$\lim_{i\to\infty} \frac{1}{b_i-a_i}\sum_{m=a_i}^{m=b_i} \delta_m(U) > 0$ by the upper Banach property of $A$.
And after choosing a further subsequence we can have a weak* limit point of $\frac{1}{b_i-a_i}\sum_{m=a_i}^{m=b_i} \delta_m$ denoted by $\mu$, and it is invariant with respect to the shift $T$.
We can then apply Furstenberg's theorem on set $U$ with $\mu(U) > 0$,
$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{i=n} \mu(U \cap T^{-i}(U) \cap T^{-2i}(U) \cap T^{-i}(U)... \cap T^{-ki}(U)) > 0$
For any $k \geq 0$.

$R = U \cap T^{-i}(U) \cap T^{-2i}(U) \cap T^{-i}(U)... \cap T^{-ki}(U)$ is then a positive measure set for some $i$, and it is not empty because it has positive measure, and every point $\omega' \in R$ will has the property that $T^{qi}(\omega') \in U$ which means $\omega'_{qi} = 1$ for $q$ any integer between 1 and $k$, now because $\omega'$ is in the trajectory closure of point $\omega$, which means any open neighbourhood will contain points of the trajectory of $\omega$, since the open neighbourhood in this case is defined by the product topology so in particular we can see this is to say that every finite segment of $\omega'$ will appear somewhere in $\omega$, so in particular we have in $\omega$ with some number $l$ such that:
$\omega_l = \omega'_i = 1, \omega_l = \omega'_i = 1, \omega_{l+i} = \omega'_{2i} = 1, ..., \omega_{l+(k-1)i} = \omega'_{ki} = 1.$
Now form the definition of point $\omega$ it is possible to see that $l, l+i, l+2i, ..., l+(k-1)i$ are all in $A$ □

A small extension:
Since the mean point in this proof is that there exist some limit invariant measure that gives the set $U$ a positive measure, so we can get a generalized form of the condition of a set of natural numbers that contains arbitrarily long arithmetic progressions:

**Corollary 2.7.** *Let $A \subset \mathbb{N}^+$, then if there exist a non-negative function $c : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{R}$ such that:*
*1.For every $N \in \mathbb{N}^+$,$\sum_{s=1}^{N} c(s, N) = 1$,*
*2.$\limsup_{N\to\infty} \sum_{s=1}^{N} c(s, N)\chi_A > 0$,*
*3.$\liminf_{N\to\infty} \sum_{s=1}^{N} |c(s, N) - c(s-1, N)| = 0$, with the understanding that $c(0, N) = 0 = c(N+1, N)$.*
*Then, in $A$ we can find arbitrarily long arithmetic progressions.*

*Proof.* We can replace the special choice of $\frac{1}{b_i - a_i} \sum_{m=a_i}^{m=b_i} \delta_m$ in the proof of Furstenberg's theorem by
$\sum_{m=1}^{m=N} c(m, N)\delta_m$, then the first property said that it is a probability measure, the second one said that we can find a subsequence $N_i$ that $\sum_{m=1}^{m=N_i} c(m, N_i)\delta_m(U) > 0$, then the third property said that among the subsequence $N_i$ we can find a subsubsequence such that the the limit measure is invariant.
Then all the steps will stay unchanged. □

The conditions of the corollary are not necessary, there are several even weaker conditions can be found. Which can make it possible to apply the theorem on the smaller set of numbers which does not have upper Banach density, for example prime numbers see [11].

So we can see that although turned down the physicist Boltzmann, recurrence theorems is still interesting and strong enough to have some application somewhere else.

The Erdös and Turán conjecture was mainly intended to show that there exist arbitrarily long arithmetic progressions among prime numbers, however the set of prime numbers does not have upper Banach density as we can see by the famous prime number theorem, that was the point of the first conjecture, the sum $\sum_p \frac{1}{p}$ diverge for the set of prime numbers but the conjecture is still unsolved.
In 2004, Green and Tao used a generalized and extended version of Szemeredi's theorem to prove there existence of arbitrarily long arithmetic progression among primes without using the Erdös and Turán conjecture, their idea is to construct another measure on numbers other than the standard counting measure (as can be seen in the corollary with the function $c(s, N)$), then proved a corresponding Furstenberg's theorem on the manipulated measure, then argue that the counting measure for prime can be majorized be the manipulated measure and showed the existence of arbitrarily long arithmetic progression among primes.

The proof of the existence of arbitrarily long arithmetic progressions in primes is not very simple but with the help of the corollary we can have some other conditions for sequence of integer to have arbitrarily long arithmetic progressions.

**Definition 2.8.** For a sequence $a_i$, we denote the corresponding difference sequence $d_i = a_{i+1} - a_i$, then define the sequence $b_i = \frac{d_i}{a_i}$

**Theorem 2.9.** *If* $\limsup_{m,i\to\infty} \frac{\frac{1}{a_i}+\frac{1}{a_{i+1}}+\frac{1}{a_{i+2}}+...+\frac{1}{a_{i+m}}}{\frac{1}{b_i}+\frac{1}{b_{i+1}}+\frac{1}{b_{i+2}}+...+\frac{1}{b_{i+m}}} > 0$*, then* $a_i$ *contains arbitrarily long arithmetic progressions.*

*Proof.* For such a sequence $a_i$, we can construct a sequence probability measure on integers which translated into the dynamical system of the shifting sequence will contain a invariant weak*-limit measure that gives us positive measure on some certain set $U$. Namely to construct a function $c : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{R}$ satisfying the conditions of the corollary 2.3.2.
Let's choose any natural numbers $m, i$, then choose the $m$-segment of $a_i$, then we can firstly assign to each number in the range $[a_i, a_{i+1})$ number $\frac{1}{a_i}$, then for the rest number we do the same thing, namely, function $c'(s) = \frac{1}{a_j}$ for $j \in [a_j, a_{j+1})$ where $j = i, i+1, ..., i+m-1$ for other numbers set $c'(s) = 0$, then we can normalize the function $c'$ by multiplying a factor $\frac{1}{\frac{1}{b_i}+\frac{1}{b_{i+1}}+\frac{1}{b_{i+2}}+...+\frac{1}{b_{i+m}}}$, where the $b_i$ sequence is defined in the

definition 2.3.3. So $c = \frac{1}{\frac{1}{b_i}+\frac{1}{b_{i+1}}+\frac{1}{b_{i+2}}+...+\frac{1}{b_{i+m}}}c'$ is a probability measure, and denote it as $c_{m,i}$.

It is then easy to check that:

$c_{m,i}(i) + |c_{m,i}(i+1) - c_{m,i}(i)| + ... + |c_{m,i}(i+m-1) - c_{m,i}(i+m)| + c_{m,i}(i+m) = \frac{1}{\frac{1}{b_i}+\frac{1}{b_{i+1}}+\frac{1}{b_{i+2}}+...+\frac{1}{b_{i+m}}}\frac{2}{a_i}$,

from the fact that $a_i$ is an infinite sequence we can actually see that the condition 3 of corollary 2.3.2 for sequence of measure $c_{m,i}$ will be satisfied but then the condition 2 in the corollary 2.3.2 will be literally the condition in theorem 2.3.4, which means for sequence satisfy the condition of the theorem there exist sequence of measure satisfy the condition of corollary 2.3.2. Which means that it will contain arbitrarily long arithmetic progressions. □

It is possible to check that actually the set of primes will not satisfy the condition of the theorem, in this sense the condition of the theorem is still too strong. But it is nevertheless an easier to check property than the condition of corollary 2.3.2(which is still not a necessary condition).

So much on number theory, so the physical uselessness of the formalism of ergodic theory or general dynamical systems can not stop us producing some more physical useless mathematical theorems. For next chapter, I will go back to talk some old issues of statistical physics, and the only argument from the first two chapters is that the Liouville equation of Hamiltonian systems is actually the Frobenius-Perron operator act on the $L^1$ space on phase space, the long detour of number theory will be not used.

# Chapter 3

# Stochastic description—some general discussions

After some detour in mathematics, I will go back on physics.
Regardless of the interpretations, the probability or stochastic description in physics is just introducing some non negative $L^1$ functions on some certain spaces and discuss the behaver of the $L^1$ under the dynamic flow, this formulation is clear in mathematics but not clear in our mind, usually we think of a distribution as some chance related concept, but here the $L^1$ functions are not actually the same of what we mean by "chance", indeed, there are still a lot of debate on the meaning of probability, here in this chapter, I will introduce the $L^1$ version of stochastic description of physics.

## 3.1 Idea of uncertainty–unnoticed but remarkable revolution in science

If i have to name a difference in mathematics between now and 200 years earlier i would say that nowadays we have a precise MATHEMATICAL theory on uncertainty,namely,probability theory.After Kolmogorov's famous booklet see [13],probability theory as a axiomatic theory grows very fast,the law of large numbers,central limit theorem,limiting properties in different cases,theory of stochastic processes...are now applied in a lot of research field other than mathematics for example statistics,economics,finance as well as physics and actually probability theory are all in the very central part of those research field.
Indeed,the stochastic way of thinking become more and more important nowadays.For example in physics,in addition to the statistical physics which i will talk about more

in detail later,the idea of quantum physics is also kind of stochastic philosophy,in the non-relativistic quantum mechanics,the wave functions has the Born probability interpretation,and the description of the real world is based on the probabilistic way,the physical quantities are random variables whose distribution is determined by the wave function and the evolution is also determined by the evolution of the wave function,in the theory if we focus on a single quantity,we can describe the quantity as a non-stationary Markov process.As for the modern version of quantum physics,the quantum field theory,the starting point is trying to give some wave function of the field,but the whole point inside is really that we have some random field whose distribution is determined by the "state",and we have a time evolution of the state derived from the operator sense Lagrange formulation,all the interesting physical quantities can be also seen as stochastic process,and the description of scattering process is based on the so called cross section which is the transition kernel of the stochastic process.Notice that the idea of probabilistic cross section is not original from quantum physics,Maxwell in one paper, see [14], discussing the heat equilibrium characterized the interaction between the gas particle and the wall as random cross section and then people in kinetic theory of gas and fluid generalized this idea to the so called scattering kernel,and use this we justify the validity of Boltzmann H-theorem as well as the entropy property of fluid material,see [15] for more and precise details.

All of a sudden,the probability theory occupied every part of science with or without being noticed by the researchers,but the word probability has no unique meaning in science,there are still different understanding on that.For example,in most cases,the probability in reality means the frequency of the appearance of some event in large number of repeated experiments,as in the sentence when throwing one dice 10000 times there are 1/6 of them has number 1 and we say the probability of the appearance of number 1 after throwing a dice is 1/6.But there are also other sentences containing probability as we always see in the weather report,the probability of raining is 0.4,that makes no sense in frequency since it is not repeatable,so the number 0.4 is just some level of confirmation.This approach is known as the epistemic interpretation of probability theory.But those interpretations make no difference in the mathematical framework of probability theory,since the mathematical probability theory is just the analysis on finite measure sets which has no meaning in any real phenomena,there is NO dynamics inside,even the theory of stochastic process is just a generalized(or specified) version of probability theory in the sense what we measure are sets of functions(paths).Law of large numbers,central limit theorem can be seen as the result of finite measure analysis,there is no dynamical relevant issue,i will discuss more on this topic later.Now i state that,the pure probability theory makes no sense in real world in describing "random" phenomena,although it seems to be really convincing and it is also very developed theory but a

lot of mathematicians and has some great application in the real world as well as in mathematics itself,but it is not quite the right way of describing the real phenomena.Actually Kolmogorov himself in the end came up with the idea of so called Kolmogorov machine to describe the "randomness" to justify the validity of his axiomatic approach to probability theory,which is very abstract and nowadays not so many people really care about that compare to the number of people who use his idea of probability theory.In this thesis i would like to give some intuitive description of the validity of several stochastic modelling in real life with some not so abstract settings.For a nice discussion of this topic see [13]

To see an example when dealing probability with physics,i will talk about the statistical physics which bare the name from probability theory.

## 3.2 Magic "distribution" in physics–From Boltzmann equation to BBKGY hierarchy

Maybe it is just a coincidence,from the last 100 years,the idea of uncertainty appears at almost the same in different branches of physics and they all lead to some new understandings about the physical world.Statistical physics appears a little bit earlier than the axiomatic probability theory but later the Gibbs approach appears after the foundations of mathematical probability theory and after that there appeared so called mathematical statistical physics with highly abstract algebraic approach within the formalism of Gibbs.Boltzmann himself,in the end,in his famous H theory,introduced the density function in $\mu$-space to characterize the state of gas system,and then study the dynamical effect of the gas induced on the change of the distribution functions,after some approximation with some probabilistic thinking,he got in the end the Boltzmann equation see [16] or [17] or [18] for precise details.

Nowadays there are still different believing on the foundations of statistical mechanics,those holds the idea against Gibbs usually says that Gibbs formalism lacks reasonable interpretation while Boltzmann's idea is clear in physical sense.But,notice that,in the end in Boltzmann's H theorem the logic or mathematical structure is actually the same as Gibbs formalism,the (probability)distribution on $\mu$-space will induce a (probability)distribution on $\Gamma$-space after some symmetry discussion of distinguishable or non-distinguishable particles,and the precise idea of constructing Boltzmann equation can be seen as the Frobenius-Perron operator induced by the Hamiltionian flow with some approximations(mainly the molecular chaos hypothesis),this idea stands now in a central position in today's study of the foundation of hydrodynamics,transporting process in materials(for example in non-ideal gases) see [17].

And as always any discussion on the discussion of Boltzmann's approach to statistical physics will mention the famous Loschmidt and Zermelo's ideas against Boltzmann see [19].But we have to admit that,in the level of philosophy this discuss would make sense because if we do not care about the precise formulation then the ideas become pure philosophical and generalized.If we just take Boltzmann's H theorem as Boltzmann's argument to the irreversibility and the entropy increasing phenomena,then neither Loschmidt nor Zermelo's critisism would apply,because the approximation in the construction of Boltzmann equation already changed the situation,Boltzmann equation of distribution function is not precise change induced by Hamiltonian flow,but if we want to say the change of distribution is induced by some dynamics it is not Hamiltonian but rather some "random" dynamics which is not invertible,and moreover the H theorem states the evolution of the value of a functional act on the function space on phase space instead of the value of a function on phase space,the behaver of individual trajectory on phase would not affect the functional directly.

Since the main point of the above discussion of Boltzmann equation is that it is not the precise equation concerning the change of distribution,so what if we make it precise and clear,indeed,there are some treatment on this topic,for example the BBKGY hierarchy (see [17] for more detail)the main idea of the BBKGY hierarchy is to write down the precise version of Boltzmann equation without the statistical assumptions,the starting point is the Liouville equation of Hamiltonian system:

$\frac{d\rho}{dt} + \sum(\dot{q_i}\frac{d\rho}{dq_i} + \dot{p_i}\frac{d\rho}{dp_i}) = 0$ where the sum is over all the particles.This is a particular case of the Frobenius-Perron operator induced by the time continuous dynamical system generated by a set of differential equations see [1] for the detail the mathematical formulation and the derivation.

Then if we specify the Hamiltonian equation,it is then possible to write down the partial differential equation explicitly and notice that in the notation of partial differential equations,the Hamiltonian trajectories are the character lines of the equation,and along the character lines the function $\rho$ is constant.Given an initial distribution(integrable and smooth) the equation is solvable in the sense there exist function $\rho(t,p,q)$ such that $\rho(0,p,q)$ is the initial distribution,and fulfil the Liouville equation,and more over the solution is unique.The construction of equation is in principle easy,for every point in $\Gamma$,we can find the full time trajectory,then we will know the point on the trajectory at time 0,then take the value of the initial distribution at that point.This method will hold also for generalized functions,in the sense that the smooth integrable functions are dense in the uniform norm in general distributions,then the solution will exist in weak sense in particular for the Dirac $\delta$ distribution,the method of character will still apply and the weak solution just represents the Hamiltonian trajectory in the sense the point the $\delta$ distribution centred tract the Hamiltonian trajectory.

Now after the brief discussion of the existence of the solution of Liouville equation let's

come back to the BBKGY hierarchy,consider the N-particle system,first we have a N-particle distribution,then write down the N-1 conditional distribution by integral out the last particle,then it is possible to determine the time evolution of the N-1 conditional distribution from the Liouville equation of the N particle distribution,then we can write down the N-2 particle distribution in terms of N-1 particle distribution we do this until we arrive the last particle,so that we have the 1 particle conditional distribution equation related to the 2 particle conditional distribution.So in the end we have a set of N integral equations,it is in general very hard to deal with but the point is:if we have particles and we know the form of interactions,then we can set up the BBKGY hierarchy equations to arbitrary number of particles and study arbitrary subsystems.With some further approximations it is possible to derive Boltzmann equation from this,see [15],and also the consideration of solution of the system of integral equations leads some interesting topics in fluid dynamics as well as granular gas,that is the main point of the kinetic theory,the study of the properties of the solutions of BBKGY hierarchy or the approximated version–the Boltzmann equation when specifying some certain form of particle interactions see [20].

## 3.3 Boltzmann's H-theorem and the objections and the objection on those objections

Before I will start I will first discuss briefly the some general formulations of kinetic theory where Boltzmann equation and H-theorem can be derived from some more general equations, then I will show that the "more general equations" are actually Liouville equations which is not very surprising, that thus it is the Frobenius-Perron operator act on the Hamiltonian system introduced in the chapter of generalized Gibbs formalism.

For kinetic theory we have some Hamiltonian systems of many particles, then we can define the $\mu$-space as the physical space with moving particles together with possible velocities and the $\Gamma$-space is the phase space which can be constructed by the product of all single particle spaces together with the corresponding topology, $\sigma$-algebra and all the structures constructed by product.
The starting point is that we have a distribution on $\Gamma$-space which we can call $F_n$ if we have n particles, a special case would be the $\delta$ distribution, then we have a certain configuration, and indeed every point of $\Gamma$-space induces a certain particle configuration on $\mu$-space, it is clear once we have a $\delta$ distribution on $\Gamma$-space then the corresponding configuration on $\mu$-space can be identified as $\sum_{i=1}^{i=n} \delta(x_i, p_i)$ but it is not very clear how to identify some general distribution on $\Gamma$-space with something of $\mu$-space, so next I will show that the distribution on $\Gamma$-space will give us multi-particle distribution in the

usual sense in physics books as:

For $0 < m < n$ a natural number, we can define :
$F_m(i_1, ..., i_m) = \int_{\mu_{j_1}\mu_{j_2}..\mu_{j_{(n-m)}}} F_n(1, 2..i_1..i_m..n)d\Gamma$.
Where for simplicity i denote the i-th particle coordinates as i and the coordinate space as $\mu_i$, and $d\Gamma$ is the Lebesgue measure, the integral runs over the not chosen particle spaces.
The formula above defines how to get particle distribution of m-tuples with given $F_n$, but we can also derive $k$ particle distribution from $m$ particle distribution whenever $k < m \leq n$.

Now let's see what's the meaning of the integrals (Of course we can argue by probability argument that the integrals give us the marginal distribution of subsystems but it will be not clear enough). For convenient I will show the formula for 1 particle distribution function $F_1$, the others are similar,

Let's denote counting function $N_A : \Gamma \to \mathbb{R}$ for measurable $A \subset \mu$ as: $N_A(\gamma) =$ the number of particles inside set $A$,
Here we don't need to distinguish or identify particles, different counting corresponds to different functions we can define.
Then it is possible to see that the integral $\int_\Gamma d\Gamma N_A F_n$ for fixed $A$ defines a functional on $F_n \in L^1(\Gamma)$, and for changing $A$ it is possible the integral defines a real values positive measure which is absolutely continuous with respect to the Lebesgue measure on $\mu$, so that by Radon-Nikodym theorem there exist almost uniquely a $F_1 \in L^1(\mu)$ such that $\int_A d\mu F_1 = \int_\Gamma d\Gamma N_A F_n$, from the definition of the $N_A$ function we can identify $F_1$ as the density on $\mu$ space induced by the $F_n$. The other $F_m$ will share similar meaning.
The Hamiltonian flow will induce a Liouville equation on the $F_n$ as:
$\frac{\partial}{\partial t} F_n = \nabla F_n \cdot V_H$,
This equation will then induce the equation of all $F_m$ for example for $F_{n-1}$:
$\frac{\partial}{\partial t} F_{n-1}(\hat{i}) = \int_i \nabla F_n \cdot V_H d\mu_i$ the others will be similar.

In principle those equations for $F_1, F_2...F_n$ will form a closed system of integral equations with the Liouville equaiton for $F_n$, and they are all determined at the same time with the evolution of $F_n$, which can be see as the Frobenius-Perron operator for Hamiltonian systems, and it is the exact evolution of the distribution on $\Gamma$-space.

### 3.3.1 m-particle distribution as the conditional expectation

From what we have discussed above we have:
$\int_A d\mu F_1 = \int_\Gamma d\Gamma N_A F_n$ which has the form of conditional expectation, if we identify the function $F_1$ on $\mu$ as a function $F_1'$ on $\Gamma$ as the constant extension on other coordinates then we can see that:
define $\mathcal{A}_1$ the smallest $\sigma$-algebra contains which makes $F_1'$ measurable, then $\mathcal{A}$=$\sigma(\mathcal{P})$ where $\mathcal{P}$ contains sets of the form $\mu \times \mu ... A \times \mu .. \mu$ where A is a measurable subset of $\mu$ and the position of $A$ in the product is arbitrary, $\mathcal{A}_1$ will be a sub-$\sigma$-algebra of the Borel algebra of $\Gamma$.
Then we have $F_1' = E(F_n|\mathcal{A}_1)$, the other $F_m$ will be similar:
$F_m' = E(F_n|\mathcal{A}_m)$,
with $\mathcal{A}_m$ the corresponding sub-$\sigma$-algebra for $m$ particle subsystems, with $\mathcal{A}_1 \subset \mathcal{A}_2 \subset ...\mathcal{A}_n$ where the last term $\mathcal{A}_n$ is the Borel algebra of $\Gamma$, the conditional expectation defined above form a finite sequence of martingale which can be seen:
$F_k' = E(F_m'|\mathcal{A}_k)$ for $k < m$.

This martingale formulation allows us to see that from smaller particle number distribution we can not in general determine the larger number particle distribution even if we have the $\delta$ distribution on $\mu$ space. So it is only possible to determine the distribution of any number of particles from the top level of $F_n$ but not from the bottom level of $F_1$. However one thing is clear that if we have some one particle distribution $f$ then we can find all possible $F_n$ which can induce $F_1 = f$, and although we have multiple choices, the Liouville equation on all the $F_n$ will gives the same evolution of $F_1$ although they are in general different, this follows from the linearity of Liouville equation and the linearity of the conditional expectations.

### 3.3.2 Approximated density

Often, in the discussion of gas or fluid dynamics we introduce the density of mass, velocity and so on, this can be seen as approximated distribution from another distribution as follows:

Let $f$ be a distribution on $\mu$-space, then let $\mathcal{A}$ is a sub-$\sigma$-algebra of the Borel algebra on $\mu$-space, then we can define the approximated distribution $f' = E(f|\mathcal{A})$.

A special case will be:

In particular is the $\mu$-space is $\sigma$-totally bounded then we can construct a $\sigma$-partition with sets of diameter no greater than $\epsilon$, and the conditional expectation $f' = E(f|\mathcal{A})$ will be:
$f'(x) = \frac{1}{\lambda(B_i)} \int_{B_i} f$, where $B_i$ is the set containing $x$, which can be seen as the coarse grained average of $f$.

The original distribution can be a sum of $\delta$-distributions centred at the positions and the velocities of the particles. Then the coarse grained density can be also approximated uniformly by a smooth function $f''$ and the $f''$ is the final smooth density approximation of the particle distribution. The smaller we chose the radius $\epsilon$ of balls the nearer the uniform distance between $f$ and $f''$.

### 3.3.3 Non-commutativity between the coarse graining and the Liouville equation

Once we have the coarse grained function $f''$ of $f$ we can then find the corresponding $F_n''$ and $F_n$, which will not necessarily unique but they will give us the evolution of the distribution on $\mu$-space induced by the Liouville equation.

Since Liouville equation is linear we can write down the equation for the difference between the two distributions:
$\frac{\partial}{\partial t}(F_n - F_n'') = \nabla(F_n - F_n'') \cdot V_H$, we can integrate this equation to get a equation for one particle distribution function but it is not necessary, directly from this we can see that however small we make our approximation between $f''$ and $f$ the difference will not be controlled in the sense that it will grow as a linear order of $t$ multiplied by the $\nabla(F_n - F_n'')$ locally at least for small time. The error can be controlled if and only if in the case that $F_n - F_n''$ is invariant with respect to Liouville equation, which means as a signed measure it should be invariant under the Hamiltonian flow, but since we are in the situation discussing the non-equilibrium the initial distribution we do not choose to be invariant, then the error will grow with time at least in the linear order with $t$ locally, and after a long time we can not justify that our approximated density will be still a good approximation.

### 3.3.4 Objection of the objection by Zermelo and Loschmidt

Finally, we can discussion the objections by Zermelo and Loschmidt. The derivation of Boltzmann equation can be done using the formulation of the integral equations directly, and the key point is to introduce some approximations, then although the Liouville equation is a exact evolution of distribution on Γ-space, the Boltzmann equation on one particle space $\mu$ is a approximated density, then the evolutions of the two distributions even treated exactly will not be the same, and notice that Boltzmann equation is a further approximation of the exact Liouville equation, so we made an approximation on density and then another one on evolution, the first approximation will be already not suitable after long time not to mention the second one. So the thing we can do is to set up some parameters and calculate how long it will be if our approximation fails up to some certain order, then within this time scale, Boltzmann equation will be a good approximation and all the discussion of the H-theory and the entropy increasing will be valid, but since the grow of the error depend also on the initial condition (distribution) which will make the calculation not practical, but nevertheless we can argue that there is no contradiction here.

Here the first density approximation is to object the Zermelo's objection on recurrence, and the second approximation (in the Boltzmann equation itself) is to object the Loschmidt's objection on invertibility. It is not that the physics is wrong, we must be clear about what we are arguing, we can not be confused by approximated theory and the exact one.

### 3.3.5 some words on the time scale

As discussed above, the time scale for Boltzmann equation to be valid is in principle checkable but difficult, now with some practical observation (experiments), we can see that although Boltzmann equation is talking the asymptotic behaver of the system after long time, but the time can not be too long, so the equilibrium phenomena can be regarded as some intermediate time scale, which is much longer than the time scale for the micro process and much shorter than the recurrence time.

## 3.4 interpretations of probability in physics

In the above section,the distribution function played an important role without being well interpreted.The question is the same with the Gibbs formalism that in real life we do not have any distributions,we do not have a set of dynamical systems with different initial conditions and our observation of any quantity is not the average of the ensembles,the notion of ergodicity does connect the ensemble average with the time average which seems to be more realistic,but notice that in real life we do not have infinite time average,one way to deal with that is to say that we do not have infinite time average but also no precise time point value,namely whatever we measure the results are always some finite time average,when this averaging time is long enough such that it will not differ from the infinite average then all the arguments still works.But the problem is now that in real life we do observe non-equilibrium phenomena,we can actually see the diffusion of the ink in water and macroscopic thermodynamic processes.With all these problems there are a lot of critism on this approach to foundation of statistical physics,see [21] for example.There is some idea to overcome this difficulty see [22] by introducing the typicality of the system.In the space of initial conditions,for example in the situation of ink in water,the initial condition which will lead to spreading the ink uniformly everywhere is typical while those lead to strange behaver is atypical,and the typical initial conditions is significantly larger than the atypical ones when we introducing some relevant measure to quantify this.But all those discussion will not have an end until we have a precise understanding on the probability in real life.

There is one understanding which is very straightforward,just believe the world is truly random in the basic level(namely,believe the God actually is playing dices),and the deterministic phenomena we observed is because after composing each single randomness into a macroscopic system the limiting distribution of some physical quantity will goes to $\delta$ distribution(think about the law of large numbers and the central limit theorem),means we have almost surely to get some certain values of some quantity when measuring.If i specify this argument in some special frameworks dealing with the outcome of the measuring and the random nature of the world i will get the Copenhagen interpretation of quantum physics,so indeed we see that the belief of the truly random nature leads to some formulations which gives us some new understandings of the world and following this ideas we observe indeed more from the world in a very fundamental level of basic particles.

Another understanding is not very common,that is we still have a deterministic world,and the random behaver in real life is just because the system is very complex and gives the pseudo-random numbers.I will formulate this idea a bit more precisely later but now let's focus on the philosophical level of this.Actually Reynolds in his paper on turbulence see [23] he pointed out(guessed or conjectured) that,each turbulence behaver is a solution

of Navier-Stokes equation,and in the real life we observe the seems random turbulence because the the fluid described by Navier-Stokes equation is unstable in the sense it is sensitive to initial conditions,since we can not control precisely the initial conditions in the end our non-capability will give us in real life a random field of the turbulence.So here Reynold's idea is still deterministic and put the reason of uncertainty to the non-capability of our access to the initial conditions.

In this thesis i would like to formulate the second idea more,and actually that is also my personal point of view.

# Chapter 4

# stochastic modeling

In the former section on Stochastic description,a general discussion was presented on probability theory in statistical physics,well i didn't give any reason why those "distribution" functions really works in describing anything,i just gave a objective statement on the thing that regardless of how we make the interpretation of the formalism (Boltzmann or Gibbs) the fact that all those calculus used in treating different problems related to statistical physics use the same axiomatic structure. And those different ideas on probability theory differs only in the sense of language,namely,how we translate the axiomatic structure of finite measure analysis into real life,different translations lead to different understanding and field of application,i will give a short discussion on some interpretations of probability in the next section.
In this chapter i will introduce a dynamical point of view of probability and the justification of Monte-Carlo method, my point here is that the Kolmogorov's axiomatic probability theory is not relevant in real life problem, we need some additional arguments to justify what we are doing by using probability theory.
This was not my personal original idea however see:

*"This theory [K's 1933 set theoretic axiomatic approach] was so successful, that the problem of finding the basis of real applications of the results of the mathematical theory of probability became rather secondary to many investigators. ..[However] the basis for the applicability of the results of the mathematical theory of probability to real 'random' phenomena must depend in some form on the frequency concept of probability , the unavoidable nature of which has been established by von Mises in a spirited manner."— Kolmogorov*

## 4.1 different interpretations on probability

Let's imagine,if we have a clear understanding of probability and we know exactly the field of application then there is no confusion at all wherever we have a probabilistic theory of anything(for example quantum physics or statistical physics).But unfortunately we do not have such clear understanding on probability,wherever there is probability there is high level confusion and endless debating and different kinds of disagreement.Above statement also works in probability theory itself.
A brief discussion on the interpretations of probability can be found in [24] at the level of philosophy.
Along the history,there are some major interpretations of probability, the classical interpretation, the frequency interpretation, the logical interpretation, the subjective interpretation, the propensity interpretation and so on.I will not stay too much time on the philosophical discussions,the main point here is that nowadays we do not have a clear understanding about probability.
I will discuss more about the classical interpretation and the frequency interpretation.

### 4.1.1 The classical interpretation

The main point of the classical interpretation of probability is that if we focus on some "random" events, then we can divide the events into the same kind of subevents which are equally undecided,then we can count the number of subevents of every event and get a number(actually a rational number in [0,1]) by dividing the total number of subevents. Notice that this interpretation only make sense when we are dealing with finite sets of events,however it is possible to generalize this idea using the nowadays standard and solid logic structure,namely,the notion of measure.Then we can see from here the Kolmogorov's axiomatic foundation is a generalization of this classical probability interpretation.
So,the classical interpretation is all about dividing the events and counting numbers.And the improved version from Kolmogorov doesn't change the situation so much.
In physics,the derivation of Bose-Einstein statistics can be seen as some counting procedure,also the idea of Boltzmann equation and the Gibbs formalism namely by introducing some distribution functions on the phase space is just introducing a certain probability measure on the phase space.And if we go on with the Gibbs formalism and follow all the way of thermodynamic limit based on that,we can see that the thermodynamic limit(quantum or classical) is just understanding the measure induced by the form of interactions(or energy functions) on phase space of systems with some parameters,then limit those parameters in some certain ways and keep track on some certain

functions(random variables) ,basically there are only first order moments (the average) under consideration,sometime there are also second order moments under consideration,then from the parameter dependence of the quantities we can formulate phase transitions.A famous and concrete example is the Ising model which is in almost every book about statistical physics nowadays.
There is then a very annoying problem of this interpretation of probability,that is,why those counting procedures which have nothing to do in real life can be proper in describing the real life? And actually,if we place this question into the situation of the Gibbs formalism of statistical physics we get the usual version of the critism:In the real life we have one system instead of ensembles,so how can the counting procedure on ensembles give us a proper description of real physics phenomena?
Notice that usually Kolmogorov's calculus for probability is regarded as a universal rule for all the interpretations,however after taking some insight what is this formulation really is,I will want to put it into the classical interpretation with some sense extension,because they share one main idea in common:the counting.

### 4.1.2 Simple version of Law of large numbers and central limit theorem

From Kolmogorov's settings for probability theory it is possible to prove the law of large numbers and central limit theorem as well as some other limiting problem under some very general conditions.However the basic idea is really not different from what I will explain as an example.

**Definition 4.1.** Denote a set $H = \{0, 1\}$,and let $H^n$ be the space of n-sequence of 0's and 1's,the measure $\mu$ on $H$ is the normalized counting measure,i.e.,$\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2}$ and the measure of other subsets is clear form the additivity(notice that for finite set case the countable additivity and the finite additivity are the same),the measure $\mu^n$ on $H^n$ is defined as product measure.

Although here I use a very simple special set $H$ but actually all those definitions works well for any finite measure sets,the logic has no difference at all.
Now let's start counting,first let's focus on the "average" of a n-sequence.

**Definition 4.2.** Let $\omega_n \in H^n$,then $\omega_n = (\omega^1, \omega^2 ... \omega^n)$ for $\omega^i$ 0 or 1,then the average of the sequence is $S_n(\omega_n) = \frac{1}{n} \sum_{i=1}^{i=n} \omega^i$

Now let's count the property of $S_n$,what I want to count is basically the number of points in $H^n$ which makes $S_n$ in some certain interval,then dividing by $2^n$ we have the

measure of set containing points in $S_n^{-1}((-a+\frac{1}{2}, a+\frac{1}{2}))$. It is clear that for any $\omega_n$ we have $S_n(\omega_n)$ is the number of 1's appear in the sequence $\omega_n$ divided by n.And by combinatorial consideration we have the number of points with $S_n(\omega_n) = \frac{k}{n}$ is $\frac{n!}{k!(n-k)!}$ where k is an integer in the range [0,n].
Then we have

$$\mu^n(S_n^{-1}((-a+\tfrac{1}{2}, a+\tfrac{1}{2})) = \sum_{k=[(a-\frac{1}{2})n]}^{k\leq(a+\frac{1}{2})n} \frac{n!2^{-n}}{k!(n-k)!}$$

where [.] denote the integer part of a positive real number.

Using the Stirling's formula,
$\log\Gamma(r) = (r-1/2)\log(r) - r + \log\sqrt{2\pi} + O(|r|^{-1})$
which holds for any complex $r$ with $arg(r) \in [\pi-\delta, \pi+\delta]$ for any positive $\delta$

$\log(n!) = \log(\Gamma(n+1)) =$
$(n+\frac{1}{2})\log(n+1) - (n+1) + \log\sqrt{2\pi} + O(|n|^{-1})$

$\log(k!) = \log(\Gamma(k+1)) =$
$(k+\frac{1}{2})\log(k+1) - (k+1) + \log\sqrt{2\pi} + O(|k|^{-1})$

$\log((n-k)!) = \log(\Gamma(n-k+1)) =$
$(n-k+\frac{1}{2})\log(n-k+1) - (n-k+1) + \log\sqrt{2\pi} + O(|n-k|^{-1})$

For the first term with the logarithms:

$(k+\frac{1}{2})\log(k+1) = n(\frac{k}{n}+\frac{1}{2n})\log(\frac{k}{n}+\frac{1}{n}) + n(\frac{k}{n}+\frac{1}{2n})\log(n)$

$(n-k+\frac{1}{2})\log(n-k+1) = n(1-\frac{k}{n}+\frac{1}{2n})\log(1-\frac{k}{n}+\frac{1}{n}) + n(1-\frac{k}{n}+\frac{1}{2n})\log(n)$

$(n+\frac{1}{2})\log(n+1) = n(1+\frac{1}{2n})\log(1+\frac{1}{n}) + n(1+\frac{1}{2n})\log(n)$

Now consider the Taylor expansion of $n(\frac{k}{n}+\frac{1}{2n})\log(\frac{k}{n}+\frac{1}{n})$ as a function of $\frac{1}{n}$ around 0 up to the order of $O(|n^{-2}|)$

$n(\frac{k}{n}+\frac{1}{2n})\log(\frac{k}{n}+\frac{1}{n}) =$
$n(\frac{k}{n}\log\frac{k}{n} + (\frac{log\frac{k}{n}}{2}+1)\frac{1}{n} + \frac{n}{k}\frac{1}{n^2}) + O(n^{-2})$

$n(1-\frac{k}{n}+\frac{1}{2n})\log(1-\frac{k}{n}+\frac{1}{n})=$
$n((1-\frac{k}{n})\log(1-\frac{k}{n})+(\frac{log(1-\frac{k}{n})}{2}+1)\frac{1}{n}+\frac{1}{1-\frac{k}{n}}\frac{1}{n^2})+O(n^{-2})$

$n(1+\frac{1}{2n})\log(1+\frac{1}{n})=$
$n(\frac{1}{n}-\frac{1}{n^2})+O(n^{-2})=1+O(n^{-1})$

Now with all those estimates we can write the final approximation formula for $\frac{n!2^{-n}}{k!(n-k)!}$

First $\log(\frac{n!}{k!(n-k)!})=$
$-n(\frac{k}{n}\log\frac{k}{n}+(1-\frac{k}{n})\log(1-\frac{k}{n}))-\frac{\log\frac{k}{n}(1-\frac{k}{n})}{2}+1-\log\sqrt{2\pi}-\frac{1}{2}\log n+O(n^{-1})+O(k^{-1})$

Now we have to choose some suitable $a$ to see how the distribution of the average when $n$ is large.
For this case replace a by $\frac{a}{2\sqrt{n}}$ and it would make the sum taking the terms of $k$ where $|2\sqrt{n}(\frac{k}{n}-\frac{1}{2})|\leq a$, which is not surprising because this it just the form of central limit theorem.

Now it is possible to give some reasonable estimation of:
$\frac{n!}{k!(n-k)!}=$
$e^{-n(\frac{k}{n}\log\frac{k}{n}+(1-\frac{k}{n})\log(1-\frac{k}{n}))-\frac{\log\frac{k}{n}(1-\frac{k}{n})}{2}+1-\log\sqrt{2\pi}-\frac{1}{2}\log n+O(n^{-1})+O(k^{-1})}$

First we don't need to care about those constant terms, which in the end can be restored by the normalization, now the $-\frac{1}{2}\log n$ term gives a $\frac{1}{\sqrt{n}}$ factor, the $\frac{\log\frac{k}{n}(1-\frac{k}{n})}{2}$ term will give a $\frac{1}{\sqrt{(\frac{k}{n})(1-\frac{k}{n})}}$ factor but in the range of our $k$ we have:
$\frac{1}{\sqrt{(\frac{k}{n})(1-\frac{k}{n})}}=2+O(n^{-1})$ with some universal O(.) for those $k's$.
Then it is easy to see that for our range of $k$,
$O(n^{-1})+O(k^{-1})$=$O(n^{-1})$.
And $e^{O(n^{-1})}=1+O(n^{-1})$.

For the rest term,
$-n(\frac{k}{n}\log\frac{k}{n}+(1-\frac{k}{n})\log(1-\frac{k}{n}))$=$n\log 2-n(2\theta^2+O(\theta^4))$,
where $\theta=\frac{1}{2}-\frac{k}{n}$ and the power series can be derived from Taylor series for $\theta$ around 0.
From our range of $k$ we can see that the range of $\theta$ is $|\theta|\leq\frac{a}{2\sqrt{n}}$,
So in the end we have for this term,
$n\log 2-n(2\theta^2+O(\theta^4))$=$n\log 2-2n\theta^2+O(n^{-3})$.

Collect all the results above we have:
$\frac{n!}{k!(n-k)!}=Cn^{-\frac{1}{2}}(2+O(n^{-1}))(1+O(n^{-1}))(1+O(n^{-3}))e^{n\log 2-2n\theta^2}$

So $\frac{1}{2^n}\frac{n!}{k!(n-k)!}=\frac{n!}{k!(n-k)!}=Cn^{-\frac{1}{2}}(2+O(n^{-1}))(1+O(n^{-1}))(1+O(n^{-3}))e^{-2n\theta^2}=$
$2Cn^{-\frac{1}{2}}e^{-2n\theta^2}+O(n^{-1.5})$
Here all the O(.) notation is universal for the range of $k$, so we can just take them into the sum of $k$ in that range and get:

$\sum_{k=[(\frac{a}{2\sqrt{n}}-\frac{1}{2})n]}^{k\leq(\frac{a}{2\sqrt{n}}+\frac{1}{2})n}\frac{n!2^{-n}}{k!(n-k)!}=\sum_{k=[(\frac{a}{2\sqrt{n}}-\frac{1}{2})n]}^{k\leq(\frac{a}{2\sqrt{n}}+\frac{1}{2})n}2Cn^{-\frac{1}{2}}e^{-2n\theta^2}+O(n^{-1}).$

Now we recall that:$\theta=\frac{1}{2}-\frac{k}{n}$,
So $n\theta^2=(\sqrt{n}(\frac{1}{2}-\frac{k}{n}))^2=\kappa_n^2$ where $\kappa_k=\sqrt{n}(\frac{1}{2}-\frac{k}{n})$.
Now let's notice that as our $k$ changes by 1, $\kappa_k$ will change by $\frac{1}{\sqrt{n}}$.
Then sum range of our $k$ induce the sum range of $\kappa_k\in[-\frac{a}{2},\frac{a}{2}]$,
So $\sum_{k=[(\frac{a}{2\sqrt{n}}-\frac{1}{2})n]}^{k\leq(\frac{a}{2\sqrt{n}}+\frac{1}{2})n}2Cn^{-\frac{1}{2}}e^{-2n\theta^2}=\sum_{\kappa_0=-\frac{a}{2},\Delta\kappa=\frac{1}{\sqrt{n}},\kappa_f=\frac{a}{2}}C'\Delta\kappa e^{-2\kappa^2}$,
Where $\kappa_0,\kappa_f$ denote the starting value and the finial value of $\kappa$, here because some integer issue the value is not exact but will not introduce higher order error than the $O(n^{-1})$ term.
Then it is possible to see that the it can be reformulate as a Riemann sum of Gaussian function, thus converge to the integral of Gaussian function, the error term goes to 0,so we see that for any $a$ we choose, the distribution will converge as $n\to\infty$ to the integral of Gaussian function, which means the distribution function will converge to normal distribution pointwise everywhere(consider the linear interpolation of the discrete case). Which means the weakly convergence to normal distribution.
This is basically the central limit theorem in probability theory with some simplification,but in general the idea of central limit theorem in probability theorem is not more than that,the difference is that the exact counting method does not work for arbitrary sets,we have to introduce some finite measure on the sets.
So we can see that the probability theory is nothing more than counting,which is static,has nothing to do with real life where we have some dynamical random phenomena, like the outcome of the price of the stocks, the outcome of the result of a dropping coin ,the turbulence of the river...If we set up the probability theory,then it is not more than that we count the initial conditions for which the coin gives us Head and the initial conditions for Tail,or the initial conditions of the river such that it gives us a certain turbulence at certain time in a certain place,or these are just counting,there is no dynamics inside,then how can this method work if it is logically unrelated to what's really happening in real life?

### 4.1.3 Statistical physics, an application of the counting procedure

Now after talking about the counting procedure on the abstract 0,1 sequence, let's take a look on the formalism in statistical physics, here in order to avoid the physics issue on what the universe real is I will not take the real life physics but the idealized Hamiltonian systems.
Consider we have a Hamiltonian system of some particles in a closed space, so for each single particle we allow it moving inside some region $\mathbb{R}^3$ closed by some closed boundary, for this we have to introduce some interaction between the particles and the boundary and we have to assume the boundary is not very strange, discussions on the shape of the boundary and the interaction on the boundary will be tedious and not very related to the discussion here, so let's simply take a simple boundary with certain interaction to make the situation clear although it is not the important point and there are many ways to do that:

For $B \subset \mathbb{R}^3 = \{x||x|^2 < 1\}$ the unit ball with boundary
$\partial B = \{x||x|^2 = 1\}$ the unit sphere.
Now let's take functions $f_\delta : [0,1] \to \mathbb{R}$ for arbitrarily chosen positive number $\delta \in (0,1)$ with the properties that:
1.$f_\delta(r) = 0$ for $r \in [0, 1-\delta]$.
2.$f_\delta(r) \geq 0$ for $r \in (1-\delta, 1)$.
3.$f_\delta$ is smooth and tends to $\infty$ as $r \to 1$.

Such $f_\delta$ exist for any number $\delta \in (0,1)$ and moreover we can manipulate the rate of the increasing to be equal or faster than exponentially increase for $r > 1-\delta$ means that we can let:

4.$f_\delta(r) \geq e^{Mr}$ for $r \in (1-\delta^-, 1)$ for any number $\delta^- \in (0,1)$ with $\delta^- < \delta$) and for any positive number $M$.

How take any Hamiltonian system and constrain the particles inside $B$ and take a new Hamiltonian $H_\delta = H + \sum_{i=1}^{i=n} f_\delta(r_i)$, where the $i$ labels the particles and $r_i$ is the distance of the particle from the origin, the new Hamiltonian system we got is almost the same Hamiltonian system except that it is constrained inside the unit ball if the initial particles are all inside the unit ball.

Apart from this we can just take the particle reflection boundary but it is then not a Hamiltonian system.

Now let's take the interactive boundary, and our Hamiltonian system corresponds to system of identical particles.

Situation 1:
We allow to put initial condition as particles can appear everywhere with bounded speed, this can be formulated as a Lebesgue measure $\lambda$ on $B^{3n} \times (MB)^{3n}$ where $MB$ is the ball with radius $M$.
Now we first choose some time $t > 0$ and count the number of particle inside $A \subset B$ where $A$ is a Lebesgue measurable set.
This can be explicitly seen as follows:
If there are $m$ particles inside $A$ then the corresponding phase space is inside one of the $\frac{n!}{m!(n-m)!}$ possible sets of the form $A^m \times (A^c)^{n-m} \times (MB)^{3n}$ after reordering of the product which is because we have identical particles.Let's simply denote the union of the sets as $C_m$ which contains exactly the phase space points such that there are $m$ particle inside $A$.
Now notice that the Lebesgue measure is invariant under the Hamiltonian flow(here is why I do not want to use the reflection boundary because then we need some additional arguments to use this result). The Lebesgue measure of the set of initial conditions which produce the situation that at time $t > 0$ there are $m$ particles inside $A$ is the same of the Lebesgue measure of $C_m$.
Now let's take the situation that the Lebesgue measure is normalized and then on the single particle space $\lambda(A) = \frac{1}{2}$, which makes $\lambda(C_m) = \frac{n!2^{-n}}{m!(n-m)!}$, now using what have done before in the last section we can see that as $n$ is large enough the measure of the set $\bigcup_{|\frac{m}{n}-\frac{1}{2}|>\frac{1}{\sqrt{n}}} C_m$ converge to the normal Gaussian integral with lower bound $-1$ and upper bound 1, and $\bigcup_{|\frac{m}{n}-\frac{1}{2}|>\delta} C_m$ will converge to the normal Gaussian integral with lower bound $-\sqrt{n}\delta$ and upper bound $\sqrt{n}\delta$ which will be arbitrarily closed to 1 if our $n$ is large enough.

Generalizations of situation 1:
For any time $t > 0$, the set of initial conditions that will induce at time $t$ the number of particles inside $A$ is nearly $\frac{1}{2}$ can be arbitrarily closed to 1 for large $n$.
Generalization 1:
We don't need to specify that our $A$ has measure $\frac{1}{2}$ similar arguments will show that we can choose arbitrarily measurable set $A$, and the number of particles at some fixed $t > 0$ will be nearly $n\lambda(A)$.
Generalization 2:
We can actually make some finite measurable partition on unit ball $B$, then the measure of sets of the initial conditions which induce at time $t > 0$ the numbers of particles in

each set is nearly proportion to its measure can be arbitrarily closed to 1 for large $n$. This can be seen as for each set in the partition we can find some $n$ and exclude a small set with measure smaller than some fixed $\epsilon > 0$, since we have finite partition say there are $q$ sets in the partition, then we can just take $n$ as the largest $n$ appeared in each argument and the set in the end we excluded has at most measure $q\epsilon$, we can choose $q\epsilon$ we small as we want. And the set in the end left has at least measure $1 - q\epsilon$ which can be arbitrarily closed to 1.

Generalization 3:

We can choose a finite sequence of time $t_1, t_2, ...t_p$ and a fixed finite partition, then similar arguments will lead us to that the measure of set of the initial conditions which will introduce at all time in $t_1, t_2, ...t_p$ a nearly uniformly distribution of the particles in the ball $B$.

Generalization 4:

We do not need to use the Lebesgue measure on the initial conditions, any measure which is tight with respect to Lebesgue measure will work, namely for measure $\mu$ with the property that $\lambda(A) > 1 - \epsilon \rightarrow \mu(a) > 1 - \delta$ where $\delta$ is a positive number smaller than 1, and $\epsilon < \epsilon(\delta)$ where $\epsilon(\delta)$ is a positive number determined by $\delta$.

Conclusions of situation 1:

In our situation, we can see that if we focus on how the particles are distributed we can get the conclusion that for large $n$, if we normalize the Lebesgue measure, a large set with respect to the measure will give us nearly uniformly distribution of the particles at any fixed time $t > 0$ or finite sequence of time $t_1, t_2...t_p$, and we can choose the sequence of time with small increments and see that for those time our $B$ is filled almost uniformly particles.

Non-conclusions of situation 1:

1.We can not say that we have nearly uniformly distribution at all the times.

2.We can not say that if we make experiments with some certain initial conditions(which we do not know precisely), will behave like the majority with respect to Lebesgue measure.

3.We have nothing to track any single initial condition in this method.

Problems of the situation 1:

Notice that this argument will work for all kinds of Hamiltonian of identical particles, the interaction of the particles will not matter so much, so we have a very unrealistic result if we conclude that the uniformly distributed nature is typical for all kinds of interactions, because we may consider a system for particles have very strong contractions, so we may imagine that all the particles will soon be all together as a cluster of

particles which will certainly not uniformly distributed, but that situation is actually not the majority if we consider the Lebesgue measure because for particles forming clusters we have that the potential energy of the system is low, then we may look in the phase space $B^{3n} \times (MB)^{3n}$ the set with certain potential energy, which will give us potential energy level sets, as our number of particles is large, the set for low potential energy is not the majority. To formulate this we need to know the certain form of interaction and calculate the potential energy level sets, this will be not easy to perform, but if we accept this we will see that the argument above has no mathematical problem but in physics it can not give us interesting information. It is like the argument in the counting of 0,1 sequences and say that if we consider all the 0,1 sequences as equally possible to appear then most of them will give us the behaver of law of large numbers and central limit theorem, but no informations will be given for any certain sequence in the counting argument.

Situation 2:
Now we are not allowed to arbitrarily put our particles, we have to choose an energy first and manipulate the initial condition such that it will have that certain energy. In doing this we actually fix an energy surface and choose initial condition from the surface, without any other constrains we can formulate the initial distribution as uniform on the energy surface which will correspond to the micro-canonical measure.

Now we are interested again how the particles are distributed in the ball $B$, but different from situation 1 we are now focus on how close the particles are in the sense that if they all form a cluster or not. This can be formulated by a phase space function as follows:
Let function $G : B^{3n} \times \mathbb{R}^{3n} \cap \{H(\omega) = E\} \to \mathbb{R}^+$ ,where $\omega \in B^{3n} \times \mathbb{R}^{3n}$ and $H$ is the Hamiltonian function, be defined as the maximum distance between the particles:
$G(\omega) = Max\{i, j||r_i - r_j|\}$ where $i, j$ run over all pairs of the labels of the particles and $r_i$ now is the position of the particle.

Now it is clear that if $G(\omega) < \epsilon$ for some $\epsilon > 0$ then the particles will form a cluster of size $\epsilon$. It is easy to see that the function is measurable as a phase space function with respect to the Lebesgue measure, it is also measurable as a function restricted on any energy surface with the induced surface measure (can be taken as Minkowski geometric content or set up the subset topology then induce the Borel measure). This argument is to make sure it is possible to talk about the measure of the set $\{G < \epsilon\}$.

Now again fix time $t > 0$ and consider the set
$C_\epsilon = B^{3n} \times \mathbb{R}^{3n} \cap \{H(\omega) = E\} \cap \{G < \epsilon\}$ which will be the set on the energy surface

with particles forming a cluster of size smaller than $\epsilon$.
Now notice that the Hamiltonian flow will make the micro-canonical measure invariant, for the measure of $C_\epsilon$ will be the measure of the set of initial conditions with at time $t$ form a cluster with size smaller than $\epsilon$.

Here we can see that this situation is much more complicated than situation 1, we have to know the energy surface according to the interactions and we have two parameters to manipulate the energy $E$ and the cluster size $\epsilon$, we are interested in the large number of particles case where there is sill a freedom to choose our $E$ and $\epsilon$ dependence on $n$. Indeed we are now in the formulation of thermodynamic limit, although it is practically impossible to make thermodynamic limit with our the measure of the set $C_\epsilon$ but we can nevertheless discuss the possible but not necessary true outcome:

For some $\epsilon$ our measure of the set $C_\epsilon$ converge to either 0 or 1 depends on the choice of the $n$ dependence of $E$, then we list the relation between the $n$ dependence of $E$ and the corresponding results of the measure of the set $C_\epsilon$ then we can see that either our system will form a cluster or not depends on the choice of our energy surface.

Conclusions of that certain outcome:
1.We can see that depends on the choice of energy surface, the majority behaver of the system is either forming a cluster or not, in this sense we can see that our system has a phase transition.
2.Again we can choose finite sequence of time $t_1, t_2...t_p$ and argue that we can have the same majority behaver at all those time, now we can notice that the $G$ function we introduced is actually $t$ continuous, so we can see that the majority of the system will either form a cluster with some size with some possible fluctuation or we have no small cluster in some compact interval of time which can be chosen arbitrary long.

Non-conclusions of the situation:
1.Again we do not have anything we say about the behaver of any single case.
2.We can not get from this a dynamical image that we can manually change the parameter and cause a phase transition of the system, because arguments here are just static counting arguments.

Problems of this situation:
1.Our situation here is too complicated that makes it impossible to get any result, there are too much parameters to be specified and the energy surface will be in general very

complicated, although in some situations we can replace the micro-canonical measure to the Gibbs-measure and the thermodynamic limit will not change and the calculation will be simpler and possible to perform by hand(for example the Ising model).
2.Although this situation gives us more interesting result if we consider it as a model of phase transition in physics, but still all the problems we faced in the situation 1 are equally existing in the situation 2, all the arguments in the situation 2 depend on the particular choice of the measure, and it is logically not different from the first situation which gave us unreasonable results, so logically we can not consider that the situation 2 will give us any reasonable results.

So if we do not care about the interpretations of statistical physics in any formalism, and we only care the logical and mathematical structure being used we can see that it is not really different than the counting method of the 0,1 sequences introduced before, although we could have very complicated formulas in thermodynamic limit and we have very interesting result which seems very reasonable in physics, but still, are those methods just coincidentally giving us reasonable answer or there is a deeper reason that all those arguments are valid.

Conclusions for the two discussions above:

Let's take a general dynamical system, or for it easier to imagine we can take a Hamiltonian system, or even simpler, identical particles with certain interactions moving inside a box, now if we construct a real valued phase space function then we can perform the level set decomposition of the phase space, the function do not need to be invariant, but it still can be almost invariant under the thermodynamic limit in the sense that at almost all the phase space point the value of the function will fluctuate around some constant, so in this sense we can construct different functions and set up the thermodynamic relation of the particle system, for this topic there exist some results for example [4], [25], but so far what we can say is that for most of the initial conditions under a certain measure we would have some certain thermodynamic relations to be observed, in real life we do not have infinitely many particles but we can argue that except for a small measure set of the initial conditions we would observe a nearly thermodynamic relation with some error, but practically that will be enough, however the initial condition we can not control and we do not have a unique counting method in the sense there exist multiple measure we can set up, they can behave quite differently, that's how we call some certain measures "canonical" because the behaver under those measures applied to real life gives us some reasonable results, but logically the "canonical" measures are

the same with those bad-non-canonical measures, however there is one property that all "canonical" measures have–they are invariant measure under Hamiltonian flow, but notice that even in particle systems we do not have necessarily Hamiltonian system nor to mention the field of Biological system or even financial systems which are other field that use the formalism of statistical physics, but actually the measure as a method of counting do not need to be invariant, we just fix some moment and count the initial conditions and check there evaluation that will be enough to give a justification of thermodynamic behaver of large systems, but that is not really a justification, let's translate the case of 0,1 sequences, for large sequence, most of the sequences will have the average around 0.5, so we pick up a sequence and the average will be around 0.5, of course that will not be valid, because it depends how we pick up the sequence, if we prefer the sequence with only 1's then we do not have the average around 0.5, the situation in statistical physics is more complicated, we do not even know how the initial conditions will be picked up, which is very important if we want to set up a counting method(a measure), why would some certain measures among all the possible measures are good to characterize the system, or is that to say: the behaver of the system depends on the measure we choose? That is out of question in statistical physics and those problem we are facing in every field using probability theory, we can set up the method to count but we can not justify why the method of counting is the natural one. The only thing we can do is to compare with reality the result we derive theoretically, but then to make the theory the "natural" one we need to check all the methods of counting and argue that they will not give the reasonable results we observe, so if there are more than one method of counting that can support the reality we can not decide which we choose just by our preference. And also we should be careful because if the logic structure is too "rich", then with suitable interpretations we can have a theory inside the logic structure to describe all possible phenomena (if they are not self-contradictory) translated into some suitable symbolic representations. Which is a topic in mathematical logic or model theory,see [26], [27]. But then that's a problem we are facing in all the science, how we can distinguish that our theory on something works is because we grasp the right point of the nature or we are just using a very strong logic structure that is possible to give us any kind of phenomena we want to describe. That will be a endless and disappointing question to ask, and i will not go deeper in logic, but using the probability theory in dynamics at least in Gibbs formalism makes dynamics not dynamical.

So the foundation of statistical physics can not only contain the axiomatic probability theory and all the theory of how the system behave under some special certain measures. Some dynamical related issue have to be concerned that how the statistical method can give us reasonable results. Next section I will discuss the frequency interpretation of probability which is a majority in practical applications of probability.

### 4.1.4 The frequency interpretation and the statistical method

To relate the probability theory to real life,there is one popular interpretation,which is the frequency interpretation,which says that,if we focus on a random outcome if some certain system,then if we record the results and do the statistics then the statistical frequency of a certain result will correlate the probability of that event.
This is one of the most concrete interpretations of probability theory but still there some problems, one thing is that if we want to explain some single case then there is no proper frequency at all since we do not have a large amount of data, well that is actually not a series problem since if we want to describe something with probabilistic method then we do not except of being able to predict anything in single case.
Yet another problem is that, still there is nothing related to dynamics, the only thing we know here about this interpretation is that ,random phenomena in real life is like some black box machine which gives random outcomes and this machine is such that the statistical behaver can be described by probability theory. So if we want to analysis some certain situation for example the throwing coin experiment then we have to know how is the black box machine for this system, and argue that this black box machine does give the random outcome which fulfil the statistics.
There is a even larger problem in this kind of interpretation,the frequency converge to probability after INFINITELY many results have been collected and averaged,but in principle we can not test the convergence of the frequency, so there is nothing we can see about the detail machine of some random outcome, we have to accept that what we are observing is random by assumption, and it fulfil the statistical condition.
So what we can do here is to repeat the experiment a lot of times and record the data and do statistics,this is main point of the research field statistics,where people propose and test distributions, use the historical data to make probabilistic predictions, for even set up some model for some complex phenomena.Actually, this is also what people in physics are doing, if we accept that we never observe something certainly, even in classical mechanics, the experiments related can not avoid all the errors, in quantum physics since the world in smaller scale is more sensitive to disturbing the error is not avoidable or even not negligible then we have to generalize our experiments to the testing of distributions, with the propose-testing method in statistics we can do physics, but we have to distinguish the method we are using to test the theory apart from the fundamental philosophy, for example, the throwing coin system gives us some "random" outcome, and we can propose the distribution and test, but it can not stop us believing that there is a mechanism behind the throwing coin system rather than just pure random outcome.That's one reason I do not take quantum physics the Copenhagen interpretation as necessary, it is fine to use the statistical method to describe physics but not necessary to introduce a fundamental philosophy for the method itself, the

quantum physics can not stop us from believing the determinism, there are no explicit contradiction to apply a statistical method on deterministic system as we have already seen in the previous chapters.
However this frequency point of view can not give us anything if we want to make any prediction, in the sense that for example if we are observing a sequence of 0,1 then we can do any statistics as we want, but the future behaver will be totally unpredictable as we will see in the next chapter with Turing machines. Now let's assume that we know a 0,1 sequence and we know all the mean values of all kinds of orders but we can not really get a unique sequence from the statistic aspect as we can see in the next sections concerning the dynamical systems.(For finite historical data we do not even know whether the sequence is statistically stable or not)
So if we believe the frequency interpretation then we have to admit that it is practically impossible to check the frequency properties of any phenomena without really knowing any more detail of the phenomena than just observing, there will be no logically well defined test of statistical property of a sequence if we only know a finite history, but there are some good test on the randomness of finite sequences as we will see in the next chapter.

## 4.2 Dynamical point of view in probability

In the beginning of this section,I will give a concrete dynamical system on $T^n$ the n-dimensional torus.

**Definition 4.3.** Let $S : T^n \to T^n$ defined by $S(x_1, x_2...x_n) = (x_1 + a_1, x_2 + a_2, ..., x_n + a_n) mod(1)$ with the real numbers $a_1, a_2...a_n$ relative irrational in the sense they are linear independent in the field of rational numbers mod 1.

It is easy to check that this dynamical system is measure preserving with respect to the Lebesgue measure.
Next it is also easy to see that this dynamical system is ergodic.
To see this it is enough to show that any real valued $L^1$ functions is almost everywhere constant,notice that we have a finite measure space which means $L^1$ functions are also $L^2$ functions,thus the Fourier transformation can be performed for any $L^1$ functions and get a Fourier series.

$\hat{f}(\overrightarrow{k}) = \int f e^{2\pi i \overrightarrow{k}\,\overrightarrow{x}} d\overrightarrow{x}$ where $\overrightarrow{k}$ is a vector of integers.

Let's assume that $f$ is invariant,then $\hat{f}(\overrightarrow{k}) = \hat{f}(\overrightarrow{k})e^{-2\pi i \overrightarrow{k}\,\overrightarrow{a}}$ then since the $\overrightarrow{a}$ vector by definition is composed by relative irrational numbers then for non-zero $\overrightarrow{k}$ vector $e^{-2\pi i \overrightarrow{k}\,\overrightarrow{a}}$ will never be 0 thus $\hat{f}(\overrightarrow{k}) = 0$ for non-zero $\overrightarrow{k}$ .So the only non-vanishing Fourier series coefficient is for $\overrightarrow{k} = 0$ which means the $L^1$ function $f$ will be almost everywhere constant.

Since the characteristic functions of any measurable sets are automatically $L^1$ functions,so the invariant sets are either with full measure or 0 measure.Thus the system is ergodic.

In this case we can say more about the system.

Any open ball $S_\epsilon$ under the dynamics will trace out a dense subset of $T^n$ ,which means that for any point $t$ in $T^n$ there is a open ball $S^m(S_{2\epsilon})$ contains $t$ for some m,then it is possible for any point $t$ to be a cluster point of a sequence of points on the trajectory with any starting point which means that all the trajectories are dense.

Yet we can still say more,consider the time continuous version of $S$ by defining $S_t(x_1, x_2...x_n) = (x_1 + ta_1, x_2 + ta_2, ..., x_n + ta_n)mod(1)$ then we can see that the Birkhoff theorem applied on smooth functions can be made stronger by replacing the almost everywhere initial points to actually all the innitial points,the argument will be similar to the proof of ergodicity using Fourier series.Then we can reduce the situation to the discrete case,the proof will be similar.For more detail and the non-ergodic-theoretic proof see [28].

I want to use the discrete case because it is convenient to use in the discussion of Monte-Carlo method simulation in computers.

Then with this result we can talk about the Monte-Carlo method on numerical integration,for example if we want to integrate a smooth function of [0,1] with value in real number,then we can set up the case for ergodic dynamical system on $T^1$ and numerically calculate the integral by using the ergodic average.And this method will also work for function on higher dimensional space.

### 4.2.1 An application–digits of integers in geometric sequence

Next let's consider another interesting problem related to number theory,we know every integer has representation in some integer base which we call x-digits,for example in the usual 10-based digital representation we can ask the question for the first digit of $2^n$, namely, 1,2,4,8,16,32,64,128,256...the first digits are 1,2,4,8,1,3,6,1,2...then as we give more and more terms of this geometric sequence does the digits has stable frequency?If so what's the distribution?

To answer this question we need to analyse the first digit of $2^n$ in 10-based representation.

First it is easy to see that the number of digit of any integer $n$ will be the integer part of $\log_{10} n$ since the statement that $n$ has $m$ digits is equivalent to $10^m \leq n < 10^{m+1}$ then take the logarithm on this inequality we get the claim.
Then the first digit can be analysed in the same way,the integer $n$ has the first digit $k \in \{1, 2, 3...9\}$ if and only if

$$k10^m \leq n < (k+1)10^m$$

where $m$ is the number of digits of $n$, take the logarithm we see that the number $n$ has first digit $k$ if and only if

$$\log_{10}(k) \leq \log_{10} n - m < \log_{10}(k+1)$$

Which is equivalent to

$$\log_{10}(k) \leq \{\log_{10} n\} < \log_{10}(k+1)$$

Where the $\{\}$ denotes the fraction part.
Now replace $n$ with $2^n$ we see that the first digit of $2^n$ is $k$ if and only if

$$\log_{10}(k) \leq \{n \log_{10} 2\} < \log_{10}(k+1)$$

Now let set up a dynamical system on $T$ with

$$S(x) = x + \log_{10} 2 \bmod 1.$$

Since $\log_{10} 2$ is irrational,the dynamical system as we have already see,is ergodic,which means:

$\lim_{n\to\infty} \frac{1}{n} \chi_{[\log_{10}(k), \log_{10}(k+1))}(S^n(x)) = \log_{10}(\frac{k+1}{k})$ for every $x$ we choose as a starting point in particular for $x = 0$,

$$\lim_{n\to\infty} \frac{1}{n} \chi_{[\log_{10}(k), \log_{10}(k+1))}(S^n(0)) = \log_{10}(\frac{k+1}{k})$$

Then notice that the left side of the equation is exactly the ratio of the first $n$ terms of $2^i$ with first digit $k$.Then we can see that, in the geometric sequence $2^n$ the numbers with first digit $k$ will appear at frequency $\log_{10}(\frac{k+1}{k})$

Then notice that the central fact we use is that $\log_{10} 2$ is irrational, for this reason we can see that for any two integers $b$ and $r$ larger than 1,if $\log_b r$ is irrational then the first b-based digit of $r^n$ will appear at frequency $\log_b(\frac{k+1}{k})$ for $k \in \{1, 2...b-1\}$.
Several generalizations can be made on this problem which help us to calculate the frequency of appearance of any finite sequence of first digit or the first digit in different base and so on,i will not stay too much time on that.

### 4.2.2 On the statistical behaver of the sequence generated by certain dynamical systems

Notice that the infinity average equalling to the integral can be seen as the law of large numbers.So particularly in the ergodic dynamical system on $T^n$ discussed before the fact that averaging function along the trajectory is equal to the integral with certain measure(the ergodic measure in this case is the Lebesgue measure) fit the intuition of averaging independent random variables with identical distribution although in this case the the outcomes when focusing a single trajectory are not really independent in the sense that the trajectory in fully deterministic.
Now let's proceed with a certain form of central limit theorem.
Again consider the dynamical system with phase space $T^n$ and the self-mapping induced by irrational rotation.Namely,
$S(x_1, x_2...x_n) = (x_1 + a_1, x_2 + a_2, ..., x_n + a_n) mod(1)$
We can also study the correspond time continuous version
$S_t(x_1, x_2...x_n) = (x_1 + ta_1, x_2 + ta_2, ..., x_n + ta_n) mod(1)$
In both cases,the numbers $a_1, a_2...a_n$ are linearly independent on the field of rational number mod 1.
To begin the formulation of central limit theorem we have to first formulate the finite average precisely.
Here in order to simplify the situation and use the counting version of central limit theorem for 0,1 sequence introduced before some additional settings have to be introduced.

**Definition 4.4.** A partition $\mathcal{P}$ on set $\Omega$ is a family of disjoint sets which covers $\Omega$.
A measurable partition will be understood as a partition consist only measurable sets provided on $\Omega$ measurable with respect to some $\sigma$-algebra.
A regular partition on $T^n$ will be understood as a partition consist only open or closed sets with the usual topology on $T^n$

**Definition 4.5.** A partition $\mathcal{P}$ is induced by function $f : \Omega \to \Omega'$ and partition $\mathcal{P}'$ on $\Omega'$ is the partition consist the inverse images against $f$ of sets in $\mathcal{P}'$

In this certain case, it is possible to see by using the regularity of the measure together with the Urysohn's lemma that for any measurable set $A$ the average visit time or the asymptotic visit frequency will equal to the measure for all trajectories, which is stronger than the one use only Birkhoff' theorem by eliminating the "almost everywhere".

To use this dynamical system to generate 0,1 sequence we can make a two sets partition $A_0, A_1$ on $T^n$, the requirement of the sets is that they are measurable, then the dynamical system will give us a 0,1 sequence when we start the system at some initial point.Now we are interested in the n-average of the sequence to be precise:

**Definition 4.6.** For any 0,1 sequence $K_i \in \{0,1\}$ the $n$-average induced by $K_i$ is the sequence $K_j^n$ defined as:
$K_j^n = \frac{1}{n}\sum_{i=jn+1}^{(j+1)n} K_i$

It is then possible to define the partition induced by the n-0,1 sequence:

**Definition 4.7.** There are precisely $2^n$ n-0,1 sequences and we name those sequence by $r$ which is a number from 0 to $2^n - 1$, which are also be regarded by the binary code, the the subset $C_r$ of $T^N$ (here i use the capital letter to avoid double use of $n$), is defined by:
$C_r = \{x \in T^N | S^i(x) \in A_{r_i}, i \in \{1, 2...n\}\}$, where $r_i$ is the i-th binary code of r.

It is easy to see that those sets $C_r$ form a measurable partition on $T^N$.

We can introduce the n-step jump dynamics by $S' = S^n$ since multiply by an integer can not change the linearly independence on the field of rational numbers mod 1,namely the numbers $a_1, a_2...a_N$ used in the definition of can be replaced by $na_1, na_2...na_N$ without changing the properties of the system because they are just another set of linearly independent numbers on the field on rational numbers mod 1.

Now the statistical behaver of the 0,1-sequence is possible to see by the partition of the n-0,1-sequences,in general the partition induced by the n-0,1 sequences is not equal in measure,by nonetheless it will have a certain partition in measure.

From the ergodicity of dynamical system $(T^n, S')$ we can see that:

$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{i=k-1} \chi_{C_r}(S'^i(x)) = \mu(C_r)$.

Means that when observing for a long time and we make the statistics of the appearance of each n-0,1 sequence,they will appear in their corresponding ratio determined by the measure of the corresponding set $C_r$

This is called in statistics the correlation, for example the easiest case we can calculate the asymptotic ratio of the appearance of the segment 00 or 01, they are the two points correlation, in general we can also calculate the two points correlation of points with a certain distance $k$ by grouping all the $k$-0,1 sequences into the groups determined by the first and last elements, and we know the measure so we know the ratio. And with this arguments we can calculate any correlations.
The point here is that the sequence generated by the dynamical system introduced before has stable statistical behaver, and in general it may not be statistically Markov nor statistically independent sequence of 0,1.
So there is a natural question, can we find some dynamical system that can generate sequences with the property that the appearance of all the n-0,1 sequences will have same ratio for all n.In this case it is possible to use the simple version of central limit theorem to justify that for that sequence the distribution of the average of n-0,1 sequence will converge to Gaussian distribution for large n.
It is possible to give an dynamical system that has this property, let's still consider the torus $T^n$ but concentrate on the simple case $T$, the dynamics is not the translation but an expanding map defined by $S(x) = 2x \mod 1$, it is possible to check that that dynamics introduced on $T$ is measure preserving and ergodic and moreover for equal measure partition $A_0, A_1$ the partition of level n induced by corresponding coded sets are also equal in measure, this is precisely what we want to have. But the case here is not as nice as the case before in the sense that the asymptotic ratio exists in that manner only for the so-called ergodic points which are those points with the property in the statement of Birkhoff's theorem. Well the case here is not too bad in the sense that we know that the ergodic points form a set with full measure. So there are uncountable such points. The existence is guaranteed but it is not a checkable property that we can not have a general method to show if some certain point is ergodic point or not, for example if we choose instead of the 2-times expanding map the 10-expanding map, then there is a interesting open question that whether some certain numbers for example $\pi, \gamma$ or $e$ are ergodic points or not. If there is some method to show that they are ergodic points then in their 10-based digital representation all the segments of numbers will appear infinitely many times.

I will talk about an application of that expanding dynamics in Monte-Carlo method in solving PDE with random walk.

### 4.2.3 On the diffusion process

In this subsection i will discuss a surprising result from stochastic process,the Monte-Carlo method for partial differential equations,here i will concentrate on the easy case namely the diffusion process in 1-d case.
The diffusion equation reads as:
$p_t = \frac{1}{2} D p_{xx}$
Here let's concentrate on the solution that $p$ is a probability distribution on $\mathbb{R}$ with a initial distribution $p_0$.
The Monte-Carlo method in this case reads:
Setting up the initial distribution on $\mathbb{R}$, randomly choose a position on $\mathbb{R}$ according to the initial distribution, then simulate the Brownian motion up to time $t$, do this large number of times we can get the statistical distribution of the end position of particles at time $t$, this distribution is approximately the solution distribution at time t.
This method still works for other partial differential equations see [29], those equations have one thing in common that is the solutions can be represented by transition kernel. Then we can choose the stochastic process according to the transition kernel in the Monte-Carlo method. In this easy case the transition kernel happens to be the transition kernel for Brownian motion.
However there is one difficulty, that is we can not simulate Brownian motion on the computer, nor in real life, since if we really look the real life "Brownian motion" the trajectory is not no where differentiable.So there is another issue to approximately simulate Brownian motion and check if the double approximation works.(using approximated Brownian motion to approximate the solution).Yet we have to really care about the detail on the simulation of Brownian motion, since we have no real random walk.

### 4.2.4 Approximate diffusion equation by difference equation

This section is to derive an approximation of diffusion equation by difference equation, there are some technical details involved, the reason of this step is to justify that instead of Brownian motion we can use random walk with certain scaling on the jump distance and the time interval.

To do this let's set up the random walk on $\mathbb{R}$, more precisely on a lattice with 0 and

the distance $a$ ,then the lattice can be seen as the double sided arithmetic sequence $na, n \in \mathcal{N}$. Then let's fix some time interval by $\Delta t$, then at some time $t$ the particle is in $na$, then next step the particle will chose in equal probability to jump left or right by one lattice point.

Well this stochastic model has no real life meaning,but it gives us good illustration how we derive the distribution function on lattice evolved by time.

Basically this is a numerical method of solving partial differential equations which has been studied quite deeply, here again i pick a simple case and the result will be useful later.

Firstly let's fix a initial condition on $\mathbb{R}$ which is a smooth function $p$ with $L^1$ norm 1.

Then we set up a lattice with size $a$, and pick up the value of corresponding lattice point, $p_{a,\Delta t}(na, 0) = p(na, 0)$,where $n$ runs over the integers.

The corresponding difference equation will be,

$p_{a,\Delta t}(na, m\Delta t) =$

$\frac{1}{2}p_{a,\Delta t}((n-1)a, (m-1)\Delta t) + \frac{1}{2}p_{a,\Delta t}((n+1)a, (m-1)\Delta t)$. Which is actually the master equation for random walk on the lattice with initial distribution $p_{a,\Delta t}(na, 0)$, the initial condition here is not necessary normalized but it will have finite sum.

Now let's extend the point function $p_{a,\Delta t}(x, m\Delta t)$ to step function by taking the right continuous step extension.Denote the extended functions as $p^s_{a,\Delta t}(x, m\Delta t)$.

Then we can take the Fourier transformation of $p^s_{a,\Delta t}$:

$G_{a,\Delta t}(\theta, m\Delta t) = \int p^s_{a,\Delta t}(x, m\Delta t)e^{i\theta x}dx$

$p^s_{a,\Delta t}(x, m\Delta t) = \frac{1}{2\pi}\int G_{a,\Delta t}(\theta, m\Delta t)e^{-i\theta x}d\theta$

Using the master equation:

$\frac{1}{2}p_{a,\Delta t}((n-1)a, (m-1)\Delta t) + \frac{1}{2}p_{a,\Delta t}((n+1)a, (m-1)\Delta t)$

We can see that:

$G_{a,\Delta t}(\theta, m\Delta t) = cos(\theta a)G_{a,\Delta t}(\theta, (m-1)\Delta t)$

So:

$G_{a,\Delta t}(\theta, m\Delta t) = cos^m(\theta a)G_{a,\Delta t}(\theta, 0)$

Now let $\Delta t = Da^2$ where $D$ is some positive number and let them goes to 0.

$cos(\theta a) = 1 - \frac{1}{2}\theta^2 a^2 + O(a^4)$.

Now again let $m\Delta t = t$ for some positive $t$, to be convenient we can assume that in this formula $m$ is integer.

Then

$cos^{\frac{t}{\Delta t}}(\theta a) = (1-\frac{1}{2}\theta^2 a^2+O((\theta a)^4))^{\frac{t}{\Delta t}} = (1-\frac{1}{2}\theta^2 aD\Delta t+O(\theta\Delta t^2))^{\frac{t}{\Delta t}} = e^{-\frac{1}{2}D\theta^2 t}+O(\theta\Delta t)$

So we have:

$G_{a,\Delta t}(\theta, m\Delta t) = cos^m(\theta a)G_{a,\Delta t}(\theta, 0) = G_{a,\Delta t}(\theta, 0)(e^{-\frac{1}{2}D\theta^2 t} + O(\theta\Delta t))$

Finally:

$p^s_{a,\Delta t}(x, m\Delta t) = \frac{1}{2\pi}\int G_{a,\Delta t}(\theta, m\Delta t)e^{-i\theta x}d\theta = \frac{1}{2\pi}\int (e^{-\frac{1}{2}D\theta^2 t} + O(\theta\Delta t))G_{a,\Delta t}(\theta, 0)d\theta$
Inside the integral we can see firstly the error term $O(\theta\Delta t)$ will goes to 0 when $\Delta t$ goes to 0,since the first moment of $G_{a,\Delta t}(\theta, 0)$ will be well defined because of the $L^1$ property of $p$.Then as for the first term in the integral we can see in the definition of $G_{a,\Delta t}(\theta, 0)$ that it is actually the Riemann sum of the Fourier transformation formula of $p(x, 0)$ and as $a$ goes to 0 it produce a error term of order $O(a)$ ,then
$p^s_{a,\Delta t}(x, m\Delta t) = \frac{1}{2\pi}\int (e^{-\frac{1}{2}D\theta^2 t})G(\theta, 0)d\theta + O(a) + O(\Delta t)$
Where the first term in the right side is precisely the solution of the diffusion equation $p_t = \frac{1}{2}Dp_{xx}$,we see that the approximation of diffusion equation by difference equation is valid.

### 4.2.5 Monte-Carlo method for solving difference equation

We have seen that we can approximate diffusion process by some suitable difference equation, that is to say we can use some suitable random walk to numerically solve the difference equation and get the approximate solution of diffusion equation.
Then important point here is the construction initial distribution,if this is done then the next step is just simulate random walk in some sense, here i will use the dynamical system introduced before for generating 0,1 sequence, the length of the sequence will be determined by parameter $t$ and $\Delta t$, precisely $\frac{t}{\Delta t}$ or necessarily the integer part.
But the initial distribution is not a big problem if it is compact supported, if not we can to approximate it by some compact supported function. In the compact supported case, once we fix the lattice then the distribution is point function of finite set, then we can use again the 0,1 sequence itself we determine the point, for example we can always adjust $a$ so that the support of the point distribution has $2^N$ points, then we divide the 0,1 sequence into N+m-segments, the first N code determine the initial position and the rest $m$ code s used to simulate the random walk, then we have the uniform initial distribution, for general case we can just group the $2^N$ N-0,1 codes in some way to approximate the initial distribution.Or we can also do it precisely by introducing some suitable partition on $T^n$ i will not go too much in detail here.
There is a small problem that the discrete distribution is not necessarily normalized,but it will not be a real problem because once we make our lattice size $a$ very small, the normalization factor will goes to 1, and all the estimations above is still valid by adding this normalization factor which goes to 1.
But this argument has a series problem, here I'm actually talking about the Monte-Carlo method for numerical solution, which means that we have to take into account that the non-precise nature of computation with finite storage memory, so in real life simulation we need computable or storable algorithms to generate pseudo random numbers, one

of which is not too far away from the expanding map introduced before which is linear congruential generator I don't want to discuss the algorithms generating pseudo-random numbers here, the point is we have a method to generate a certain sequence which we treat as random sequence and put into simulation solving equations, taking the linear congruential generator as an example we can see that it is actually an approximated version of the dynamical system introduced before(with some small change or composition of the two dynamical systems) and the depends on how close the approximation is we could have numerical integration method still valid for some special class of functions, as for the numerical method for solving PDE the same argument still holds.
Taking all the above discussion in to account, the Monte-Carlo method doesn't come for free using the Kolmogorov's calculus of probability, it depends on how we generate the sequence and in other words it depends on the sequence we choose, the sequence generated by dynamical system is just a very narrow special case of sequences.
The discussion on randomness as a property of the sequence itself instead of the dynamics generating the sequence is another topic in mathematical and logic, i will give a brief discussion later.

## 4.3 Random or deterministic?

After some justification of law of large numbers the central limit theory theorem and the Monte-Carlo method,we can see that the central point in all those discussion is not the randomness of the system but really some properties of the sequence we use(for example the 0,1-sequence generated by the ergodic system).For example, the 0,1 sequence we got from the ergodic system can be stored somewhere and we can use the same sequence to simulate at any time we want, the sequence is anyway deterministic.
Randomness is very a difficult concept, and it is actually hard to say if our universe has truly random phenomena, but one thing is clear, probability theory can not describe all kinds of random phenomena, for example the random sequence witch has no stable statistical behaver, then the outcome of the random sequence can not be modelled by random variable, yet it is still possible to model it by some stochastic process, however the sequence can still fail to have any stable transition property for stochastic process. So i would say the probability theory fits for some certain processes with some general but nonetheless stable statistical behaver, for example the 0,1 sequence generated by the ergodic system I introduced before. But in real life, we can model actually a lot of systems with probability theory, and we do not even know how the system exactly works, there should be some intrinsic connection between the probabilistic description with the systems itself which is maybe beyond our reach to study if we do not even know how the system works.Here i used the word "maybe" because it is still possible to

understand the system without knowing exactly the phase space and the dynamics(here i restrict myself to dynamical systems) by only knowing the statistical behaver. But if this does not work in any system then we are facing a hard question, because in this sense the probabilistic way of thinking will not be compatible with the deterministic way of thinking. But anyway now we do not have any clue if the world is truly random as in quantum theory or the randomness in quantum theory is actually a outcome of some complex mechanism which happens in a very small scale that we do not even have a chance to attack in reasonable future.

Let's do not go too much into the philosophical discussion of randomness, so far what we can do is to use mathematical logic to enclose the probabilistic method we are using in science or real life. The Kolmogorov's axiomatic probability theory can not cover all the things we need for randomness, and logically it is not more than counting numbers so some dynamical point of view will be helpful to relate the probability theory into real life, that is what I have introduced in the last sections. Yet there is a special case which is very interesting of this dynamical point of view, the algorithm point of view where we only have at most countable sets and all the relevant method are finite in some sense and actually the algorithm discussion is very popular in the field of logic and computer science but in physics it is a little bit out of question but it is a interesting topic in the study of probability theory and general random phenomena.

# Chapter 5

# Randomness—A mathematical logic point of view

Since the main part of the thesis is discussing the meaning of probability method used in physics we have to discuss at least in one moment the meaning of probability, and indeed other than some former interpretations of probability which are very philosophical without clear formulation, there is one topic in mathematical logic and nowadays mainly in computer science and information theory. Here Kolmogorov's complexity, compressibility, randomness test will be discussed and in the end there will be a short analogue to physics.

In this chapter, an algorithm point of view in randomness will be discussed. There are several different points of view here and due to the time limit I'm not able to reach them all. For a short discussion of topics in this area as well as the historical introduction see [30].
First let's read a sentence from Von Neumann which briefly showed the idea of this chapter.

*"Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin. For, as has been pointed out several times, there is no such thing as a random number — there are only methods to produce random numbers, and a strict arithmetic procedure of course is not such a method."—Von Neumann*

The randomness in the frame of algorithm information theory is another point of view in random phenomena, instead of viewing the random phenomena as some ensembles it

treats single case where we have some certain sequence at hand or we have some method to generate sequence, then how random the sequence is?

Here some theories about recursive functions, Turing machines, computability will appear without showing the properties, for details see [31], [32].

## 5.1 Kolmogorov's complexity

Let's start with Kolmogorov's complexity, intuitively speaking it deals with how many information we need to specify the sequence, so first some mathematical setting have to be clarified:

**Definition 5.1.** The alphabet of a symbolic system is a set $A$.
We can the symbolic system is finite, countable etc. if the corresponding set $A$ is finite, countable etc.

**Definition 5.2.** A (finite) sequence is a at most countable (finite) ordered arrays $a_I$ with $a_i \in A, i \in I$ where $I$ is at most countable.

**Definition 5.3.** An algorithm $A_f$ for function $f : \mathbb{N}^m \to \mathbb{N}$ which is partial recursive is a Turing machine which computes the function $f$.Where $m$ is some positive integer.

Note that for any function which is partial recursive there exist not necessarily unique Turing machines which computes it, but nonetheless the existence has no problem.
Next let's consider for simplicity the symbolic system contains 0,1 digits, then any Turing machine which coincide with a partial recursive function of one variable can be regarded as a function with input as binary codes of the integral variable, and for the inputs which gives us result we can apply the Turing machine on the result again and again, if in all the steps we have certain result then we have a iterated sequence generated by the Turing machine with a certain input.But since the iteration of Turing machine is also a Turing machine, so without loss of generality we can consider one Turing machine with certain input and output property.
For any integer $n$, let's denote $l(n)$ the length of the binary code for $n$.
Note that here the binary code is not necessarily the standard binary expansion or the binary digital representation of integers, any bijection from a subset of all 0,1 sequences to natural numbers will be called a coding method, but later we should consider only computable coding method in the sense there exist an algorithm to let us determine the

code of any integer $n$ and an algorithm to let us determine the corresponding integer for the code.

**Definition 5.4.** The Kolmogorov's complexity with respect to algorithm $A_f$ for finite sequence $a_I$ where $I = \{1, 2, 3, 4...n\}$, is defined by:
$K_{A_f}(a_I) = min\{l(x) \in \mathbb{N} | f(x) = a_I\}$ if such $x$ exist,otherwise set $K_{A_f}(a_I) = \infty$

**Definition 5.5.** We say a sequence $a_I$ with length $n$ is compressible to $n - c$ if there exist an algorithm $A_f$ and an input $x$ such that $f(x) = a_I$ where $x$ together with $A_f$ has a description with at most $n - c$ digits.

Here for compressibility we have to find some suitable algorithms but the Kolmogorov's complexity is related to one certain algorithm.Soon we will see that it doesn't matter too much.

**Theorem 5.6.** *There exists a Turing machine $A_0$ with property:*
$K_{A_0}(a_I) = K_{A_f}(a_I) + O_f(1)$ *provided that $K_{A_f}(a_I)$ is finite.*

In [30] we can find a proof for that.
This theorem is usually called the invariance theorem, which means that it doesn't matter so much for the choice of the certain algorithm we choose, and also since there is a universal Turing machine which can simulate any other Turing machine and any finite sequence as a certain number after decoding can be always regarded as a result of a Turing machine by partial recursive function theory, so we can see that the $K_{A_0}(a_I)$ gives us always finite result with an upper bound $l(a_I) + c$ where $c$ can be chosen as the length of the Turing machine that copy the input to output which is a constant of $a_I$.Also as we can prove that $K_{A_0}(a_I)$ is an increasing function for any recursive coding as the corresponding integer for $a_I$ goes to infinite with a lower bound that also goes to infinity so when we are dealing with big amount of data then the constant difference is not very important.

As for the compressibility we have to deal with different algorithms but again it will not matter too much because the existence of universal Turing machines, changing from a certain Turing machine to the universal one gives as again some constant which is independent of the input $x$, so in some situation when we know that the $K_{A_0}(a_I) > l(a_I) - c$ then we know there will be no algorithm which can compress the sequence significantly up to some shift of constant.

After all these settings finally we can starting talking about the Kolmogorov's randomness of sequences.We can set a quantity
$\delta(a_I) = |l(a_I) - K_{A_0}(a_I)|$ which tells us how much we can reduce the large sequence $a_I$ to some smaller description. When $\delta(a_I)$ is large then we know that it is highly compressible and the information we need to know to specify the sequence is much smaller than the length of the sequence, in this sense we consider the sequence as not random in the degree $\delta$.

In order to be not very abstract i will give a very easy example.

Let's consider the algorithm which compute the function $f$ as follows:
1.The domain of the function $f$ is certain finite 0,1 sequences which has the form of $0X$ or $1X$ where X is arbitrary 0,1 sequence.
2.$f(0X) = 1^{n(X)}$ where $n(X)$ is the corresponding integer of the code $X$ which means a sequence with $n(X)$ ones.
$f(1X) = X$

Now the algorithm of $f$ can be described by some code which do not depend on the input.
Then let's consider the two sequences:
1111111111111111
1001110100100111
Both have 16 digits.

Now let's using the standard binary expansion of integers.
The first sequence can be compressed by the algorithms with input 010000,while the second has to be 11001110100100111, now we may consider the length of the algorithm itself, but it will be a constant, with this example we can also work with longer sequences for example sequence with $2^{32}$ ones which can be expressed in 0100000 which contains only 7 digits while the algorithm length $c$ can be smaller than $2^{32} - 7$ significantly. If we are still not satisfied we can take $2^{2^n}$ ones as a sequence and the compressed version is $01(0^n)$ contains $n + 2$ digits.

We may confused with the special algorithm above that take some special sequences and mark them with shorter sequences, with this we can construct the algorithm to compress any sequence significantly, yes but remember we are allowed to use one one algorithm at one time, if we want to construct some mixed algorithm then the description

of the algorithm will be larger, so there is always an issue to find the smaller algorithm which is not as powerful as the larger algorithm.And by a theorem below the large compressibility is in general not possible. If we consider the compressibility to have some rate for example we want to compress the sequence by 0.9.

**Theorem 5.7.** *For sequences with length n,any positive number* $r \in (0,1)$*, fix a universal algorithm, then with this algorithm the relative ratio of the sequences that is possible to be compressed of rate* $r$ *is less than* $\frac{1}{2^{rn-2}}$

Again in [30] we can find a generalized version of this theorem.

Which means that among all the sequence of certain length, the compressible sequences with any rate are not majority.

The Kolmogorov's complexity tells us whether given a sequence we can some useful patterns inside which helps us to compress the sequence to a shorter one, and here the patterns must be computable by algorithms and we have to fix the algorithm before doing all those discussions.
A related topic is the randomness test by Martin-Löf which will be in the next section.

## 5.2 Martin-Löf randomness test

The central idea of Martin-Löf test of a sequence is to evaluate some significant test functions $\delta_m$, where the $m$ is called significant level.
For each $m$, $\delta_m$ can be seen as a proper choice of a sample of the sequence and evaluate a number according to the sample. Here there have to be some constrains on the sampling procedure, for Martin-Löf test they have to be recursive.
Then according to the value of the $\delta_m$ we can decide to reject the randomness or accept the randomness up to the significant level $m$, we have to be sure that the sequence we reject in the significant level $m$ is really special in the sense all those sequences can can be rejected in the level $m$ form only the minority among all the sequences. For example we can require that in each level $m$ the number of sequences that can be rejected is no greater than $2^{n-m}$ among all those sequences with length $n$.
In other words, performing the randomness test is to check if there are possible recursive patterns inside the sequence, and we have to be sure that for large $n$ (which we are interested in) there have to be some sequences that can pass the test otherwise the test

tells us there is no sequence which can be regarded as random. And it has positive answer, for large enough $n$ there are sequences with length $n$ that can pass the test and we can regard them as random in the sense there are not much patterns we can follow from the sequence. The proofs contains the universal test which is not very different from the construction of universal Turing machine, and some further technical calculation on the sequences space which is not very different from that in the simple version of central limit theorem.Details again see [30].

### 5.2.1 short discussion of the idea of algorithm approach to randomness

From the two sections above, we can see that the central point of the complexity and randomness is that by recursively sampling the sequence can we find any patterns that is computable. That idea is quite close to a counterpart in physics, if we consider the recursive property as the logical version of causality and determinism in physics we can see that those random phenomena can be regarded as the procedures which gives us very complicated results that to make any theoretical description we need almost as much information as the results themselves, and the compressibility of sequence can be seen as a logical version of the theoretical descriptions of the experimental results, namely, if we can find some way to compress the experimental results into shorter descriptions. Combine all those together, the randomness in algorithm approach suggest us another interpretation of random phenomena in the real life, that is, for a phenomena to be considered as random if we can not find a description of the experimental results which contains much less information of the results themselves. In other words when we make some experiments and get some results, we can not find any method to compress the data into any shorter description which can produce that data, as an example let's take the real life common coffee-milk mixing system, what we observe is that the coffee and milk mix themselves together with some patterns which is very complicated, now we capture the patterns and want to find a description or a theory to generate the exact pattern, after one second we face the problem, not to mention if is it corrected to make the classical mechanical model with massive objects in some closed space, suppose it is corrected, then we indeed have a very short description of the mechanism, but to generate the patterns of coffee-milk mixing we need to specify the initial conditions also, which contains very large amount of data or information. So without any capability of the fully description of the system we can consider the system as random or we can assume that the system itself has some random mechanism, then instead of studying single case we can study the ensembles which actually has no meaning in any single case.

## 5.3 A toy logic model of physics

To make the idea clear I would like to present a logic model in the frame of recursive functions to illustrate the situation we have in understanding physics.
We know that the set of all recursive functions can be represented by the set of Turing machines in the sense that any recursive functions we can find a Turing machine which compute exactly this function.
How suppose I choose one of the Turing machines which corresponds to a one variable recursive function say $f(n) \in \mathbb{N}$, then I tell someone else,say Bob, that there is now a recursive function which generates sequence $f(i), i \in \mathbb{N}^+$. Now Bob knows that the sequence must be determined by some Turing machine, which in physics is the causality and determinism, by just observing the sequence generated up to any finite length he can find the Turing machine(s) which gives the first segments of the sequence. He can list the Turing machines as much as he can and select from them those give the right result up to the first segment of any length, but there is a theorem which prevent him to actually find all Turing machine which gives the right result, this will be formulated as:
Consider all Turing machines with enumeration $\phi_i$, then Bob can formulate the input also as single integers,thus we can actually consider all the Turing machines are one variable functions, then he can compute the result of its first segment of length $n$ as $\phi_i(1:n)$ and compare with the data I gave to him, then he can decide to keep the machine if it gives right result or reject it.
Consider now the set $K = \{i|\phi_i(1:n) = a_{1:n}\}$ where $a_{1:n}$ is the first $n$ numbers in the sequence I gave to Bob. Unluckily he cannot determine the set $K$ by any algorithm, because $K$ is not a recursive set although as a set it is well defined.
By some method Bob can find some Turing machines which work well up to the first $n$ numbers of the sequences, then among those machines there could be one that is what I chose but not necessary, what he can do now it to use those machines to predict the results of the sequence, he can either choose one of them and consider it to be the right one and believe that is the machine I gave to him, or he can keep all the possibilities and give a distribution among them, the according to the distribution of the machines he can calculate the distribution of the next result.
For the first case it is possible to reject it in one or finite steps if the machine gives wrong result, for the second case the situation is more complicated, he can observe as much as he can and refine the distribution he put on the possible machines which match the result in the long run which is always possible.Which means in the second case he can never reject his idea on the probability descriptions of the machines, but he knows that there is one certain machine he has to find, then he can construct a new machine which in each step decide which machine to choose, then the composition of the deciding machine

together with the machines giving "right" results is a new Turing machine which gives the right result in the exact sense as well as the statistical sense, but now he will not care about the exact sense because he can always reconstruct the deciding machine to match the historical data and keep the statistical behaver.

After some very long term of observation, he can either find a single machine which works well or find several machines together with a deciding machine which works well in statistical sense, but the question is did or can he find the machine I gave to him?Sadly enough, the question is not decidable for him even if it could be true again in the sense he can not find any algorithm to tell if he got the right result or not. For me the question is clear, I only need to check his answer and compare the recursive function he found to my original function.

Now with this conclusion that Bob will never tell whether he found the right machine or not we can see the situation we have in physics. We have a much more complicated situation but the basic principle of physics which based on the experimental results and trying to find some consistent logic structure to fit the data is not much difference than the situation of Bob, and the methods of Bob actually correspond so far what we did in physics.

So let's make a conclusion, it is very clear that if we consider all possible sequence then we will not find any way to make any kind of prediction and we can not proof that our theory for the sequence is correct or not, but the arguments above told us that even if we know that the sequence is deterministic by some partial recursive function there is still no way to make any kind of prediction and test our theory is corrected or not, this is a disappointing conclusion, but still we can not see that the nature is not understandable because the situation is different than the logic toy model I gave above, however it is still possible that understanding the nature by the experimental data can not lead us to any ultimate law of the universe, and even if we found some laws which fit any experimental data we can not really test whether the law is the right one, so one thing I believe that there is no ultimate law of the nature, what we can do is to extract more and more patterns or regularities from the experimental data can compress them into some abstract "algorithm", the more we can observe the more we know from the nature, and the more we know from the nature the more experiments we can design to observe even more, so there is a positive circuit of observing-understanding-more-observing, but there is still possible that at some point our understanding can not lead us to more observation, or the new observation can not lead us to new understanding, in such a situation, any kind of logic structure that support the observation with suitable interpretation will work as a final law of nature and we will face the end of the science, or at least physics, but that will not happen because there is no way to check whether we already observe everything based on the current technology, from this point of view the situation we are facing in real life is much better than the one of desperate Bob.

# Chapter 6

# An interesting deterministic Ising-like dynamical model

In the previous chapters I mentioned that the well known famous Ising model in statistic physics is actually not dynamical, the whole treatment for that is only just counting the points in phase space with the counting method induced from a artificial Hamiltonian function. In this chapter i would introduce a dynamical model which mimic the behaver of Ising model, but the dynamical model I will introduce is not a Hamiltonian system, so that there is no energy, no temperature, no Gibbs measure. Everything is well defined deterministic dynamics, and the price I paid is that the model is not realistic in any sense but nevertheless it will be a dynamical model with some phase transition in some sense without arguing with Gibbs formalism.

## 6.1 Construction of the model

### 6.1.1 background space

First consider the $d$-dimensional integer lattice $\mathbb{Z}^d$, let $\Lambda \subset \mathbb{Z}^d$ be a "box" in the sense that $\Lambda = A_1 \times A_2 \times ... \times A_d$ where all the $A_i$ are bounded interval of integers.

Now to each point $z \in \Lambda$ we associate a finite cyclic group $G$ with $|G| = 2n$ where $n$ is an even natural number. In other words:
$G$ is isomorphic to $\mathbb{Z}/2n\mathbb{Z}$ with the representation points $\{1, 2, ...2n\}$, we can uniformly distribute the $2n$ points on a circle, so that each $\frac{1}{4}$-arc has the same number of points.

We can paint the circle by black and white in the following way:
$(0, \pi)$ is black,
$(\pi, 2\pi)$ is white.
Here with open interval will not give us any trouble if we require that the points distributed on the circle will not be in $0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$ which is always possible hence every points is either black or white.

Then the background space of the model will be $C_z$ with $z \in \Lambda$ the circles associated to each point of the box $\Lambda$.

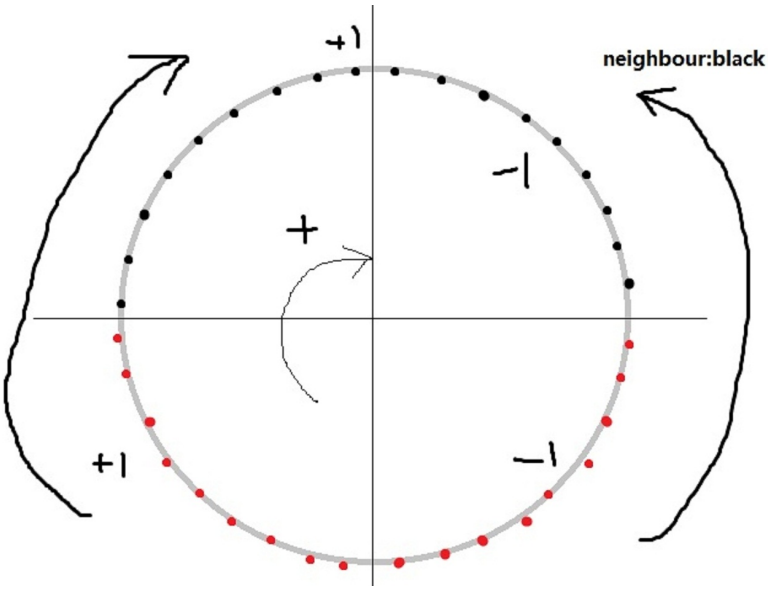
### 6.1.2 dynamics

After constructing the background space, we can put particles in each circles on one of the points, say $p$, of the circle, then for each particle we can give an "velocity" $v$ which is an integer, then we have particles on each of the circles with velocities, the dynamics goes as following:

1.For each particle, it jumps to a new point of the circle by:
$p' = p + v \pmod{2n}$.
2.For each particle, the velocity changes according to the old configuration in the following way:

The particles will have neighbours which are the particles on the adjacent circles. The velocity of one certain particle will change according to the old position of that particle and the old positions of its neighbours.
Let's denote the neighbours of particle on $C_z$ by $C_{z_i}$ where $i$ runs from 1 to the number of the neighbours.
For each $C_{z_i}$ we associate an integer field on the circle $C_z$, if the particle on $C_{z_i}$ is black(white) then the integer field on $C_z$ is such that for points in $(\frac{\pi}{2}, \frac{3\pi}{2})$ we have value $-1(+1)$, for other points we have $+1(-1)$, then we sum up all the field induced by the neighbours, and the velocity will change according to the summed up value of the point where the particle is located.
Intuitively, the velocity change can be seen as the particles want to turn the color of their neighbours to the same color as themselves.

FIGURE 6.1: Ising dynamics visualized




The visual illustration of the dynamics as above in Figure 1.

### 6.1.3 control parameter 1: the maximum velocity

As the dynamical model above there is no parameter other than initial condition we can control, so here I introduce a control parameter the maximum velocity called $v_m$, it can be seen as following:

After one step dynamics of the model, we can check whether the velocities of the particles are beyond the maximum $v_m$ or below $-v_m$, for particles with velocity outer the range we can reset the velocity back to the range by $v > v_m$ we set $v = v_m$, $v < -v_m$ we set $v = -v_m$.

### 6.1.4 control parameter 2: external force

As for Ising model it is also possible to introduce an external force here, by adding some certain number $b(-b)$ to each velocity for each step (according to the position of the particles same as the picture above), and of course we can have at the same time the external force and the maximum velocity they are not contradictory.

## 6.2 Some qualified properties

Now that we have the size of the circles (number of points on it) $2n$, the external force $b$, and the maximum velocity $v_m$ those are easy to control parameters, as for the shape and size of $\Lambda$, the initial conditions will be complicated, later in the simulations I will choose the shape of $\Lambda$ easily. But before the simulations I would like to introduce some qualified properties of the model which deal with only $2n$,$b$ and $v_m$ as well as the dimension $d$.

### 6.2.1 0 external force

Now let $b = 0$, there is no external force, then consider particles not on the boundary (we can actually round this boundary problem by setting periodic boundary conditions of the box $\Lambda$ then there will be no boundary) then it will have $2d$ neighbours, then the possible velocity change will be $2d, 2d-2, 2d-4...0, -2, -4... -2d$ depends on the configurations of the neighbourhood.
Now let's consider the case that the neighbours are frozen namely, the neighbours are fixed in color. Since the situation is symmetric in exchanging the color so we need only to consider $2d, 2d-2, 2d-4...0$ cases which means there no more white points than black points in the neighbours. The particle itself will move on its own circle and it will eventually enter the black area, depends on the velocity it enter the black area it can escape or be trapped forever, since there is a maximum velocity, so if the particle with maximum velocity can not escape, then all other velocities will also not escape, in this case we have some conditions on the maximum velocity.

For velocity change $\alpha$ which can be $2d, 2d-2, 2d-4...0$, we can exclude the case 0 since we know that all particles with any non-zero velocity will run around the circle without being trapped anywhere.
In this case for the particle with maximum velocity being trapped if and only if:
$v_m + v_m - \alpha + v_m - 2\alpha + ... + v_m - [\frac{v_m}{\alpha}]\alpha \leq \frac{n}{2} - v_m$ which can be rewritten as:
$v_m([\frac{v_m}{\alpha}] + 2) - \alpha\frac{(1+[\frac{v_m}{\alpha}])[\frac{v_m}{\alpha}]}{2} \leq \frac{n}{2}$.

For a good approximation of the the range of $v_m$ we can actually drop the integer part symbol [.], then what we get is:
$v_m^2 + (3\alpha)v_m - \alpha n \leq 0$ we can solve the inequality and get the upper bound of $v_m$: $\frac{-3\alpha+\sqrt{9(\alpha)^2+4\alpha n}}{2}$.

An illustration of this for $d = 2$ will be:
For $d = 2$ there are $4, 2, 0$ cases of the velocity change, then we have two upper bounds correspond to the case 4 and the case 2, we call them $v_4, v_3$ since the case 4 means the particle is surrounded by four black points, and the case 3 means the particle is surrounded by three black points and one white points.

If $v_m > v_4$ we can expect that the system will be not in order, there will be no stable clusters since even with four black neighbours it is still not certain that the particle can be trapped in color black.
If $v_4 > v_m > v_3$ there will be clusters with stable inner points in the sense that the inner points of the cluster will be trapped. But on the boundary of the clusters everything can happen.
If $v_m < v_3$ then we have stable clusters in the sense that some boundary points will also be trapped.

For other dimensions we can see the same results, the only thing is that for higher dimension we have more levels of the bounds, which corresponds to the stability of cluster in certain type (with certain type of boundary).

We can see here that depends on the maximum velocity we set, we can observe the system as either with no stable cluster which is very disordered, or with stable clusters which can be seen as ordered, results of the simulation will confirm this arguments which will be shown later.

### 6.2.2 non zero external force

All the arguments will be the same as 0 external force case, the only thing here is that the velocity change will be:
$2d + b, 2d - 2 + b, 2d - 4 + b...0 + b, -2 + b, -4 + b... - 2d + b$. Here I will discuss the case for $d = 1$. Other cases will be similar but more complicated

$d = 1$:
There is only one upper bound in $b = 0$ case, and the boundary can never be stable, however it is possible that for $b$ when large enough can make the boundary stably trapped in one certain color, namely,:
$v_m([\frac{v_m}{b}] + 2) - b\frac{(1+[\frac{v_m}{b}])[\frac{v_m}{b}]}{2} \leq \frac{n}{2}$.

It is a relation of $v_m$ and $b$, if the condition is fulfilled then we can observe that eventually all particles will be trapped in one certain color. If it is not fulfilled then it is still possible that the distribution of the color will prefer a certain one, but if $v_m$ is such that even for $b = 1$ the condition is fulfilled then we can observe a ferromagnetic phenomena in the sense that even with the smallest external force we will have a total ordered system.

Notice that, the arguments here is not the real case for the model, because the assumption of the frozen neighbours, however the number we derived from those arguments will be a relative good approximation to characterize the behaver of the system.

## 6.3 Simulations for the case d=2

### 6.3.1 0 external force

Here I ran some simulations of the model in $d = 2$ case, with square shaped $\Lambda$ with periodic boundary condition (torus), I chose the initial conditions (pseudo)randomly and run for some certain steps, we can see from the simulations the role of upper bounds $v_4, v_3$.

Due to the computation capability of my personal computer the size of $\Lambda$ will be no greater than 200, for the size of the circle I will choose to be $n = 500, 1000$, then the corresponding upper bounds $v_3, v_4$ will be listed:

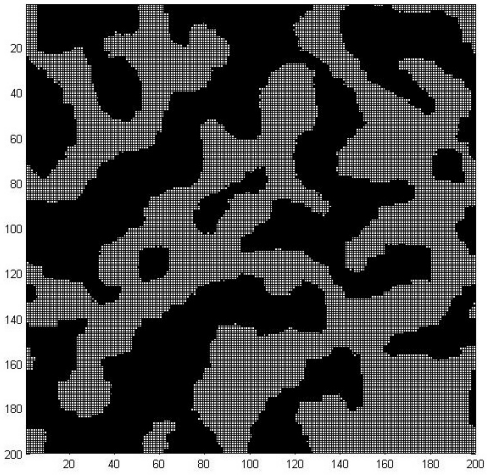
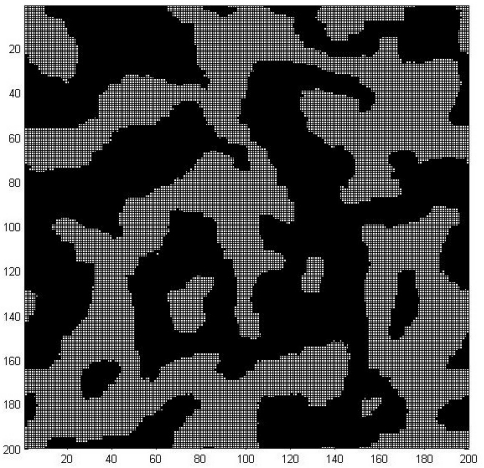
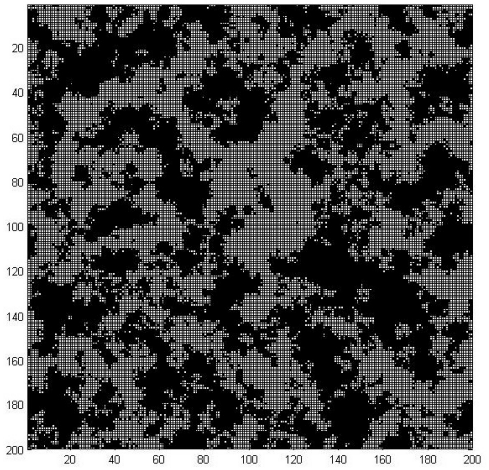
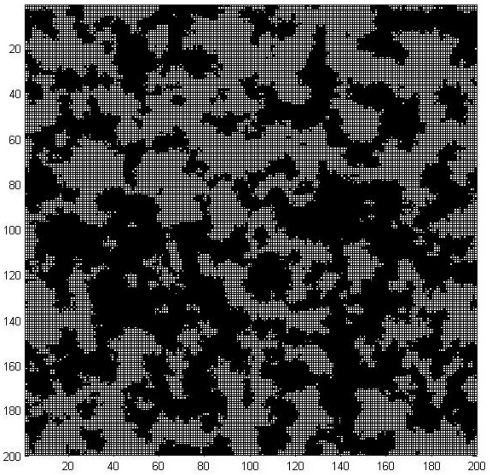
For exact value:
$n = 500$:$v_3 = 28, v_4 = 39$
$n = 1000$:$v_3 = 41, v_4 = 57$

Then as talked above, if $v_m < v_3$ we would expect stable clusters with stable boundary while for $v_m > v_4$ there will be no stable clusters, for $v_3 < v_m < v_4$ there can be stable clusters but with out stable boundary, here are the results for the different $v_m$ for each $n = 500, 1000$. In their corresponding $v_m$, systems with different $n$ will show similar properties.

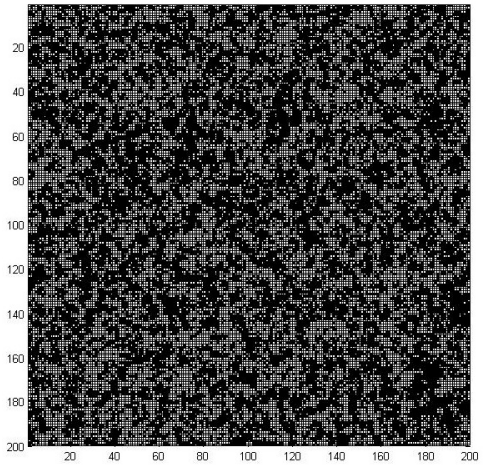
From the pictures, the role of $v_3$ is clear since we can observe really huge difference of the stability and the size of the clusters and their boundaries but the role of $v_4$ is not

quite clear, but if we observe in detail we can see that actually if $v_m < v_4$ then the system has clusters but maybe not with stable boundaries.

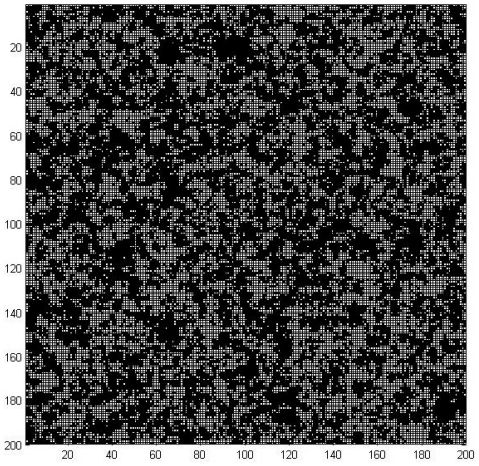
TABLE 6.1: 0 external force simulation

| | |
|---|---|
| n=500,$v_m$=25 | n=1000,$v_m$=40 |
|  |  |
| n=500,$v_m$=32 | n=1000,$v_m$=45 |
|  |  |

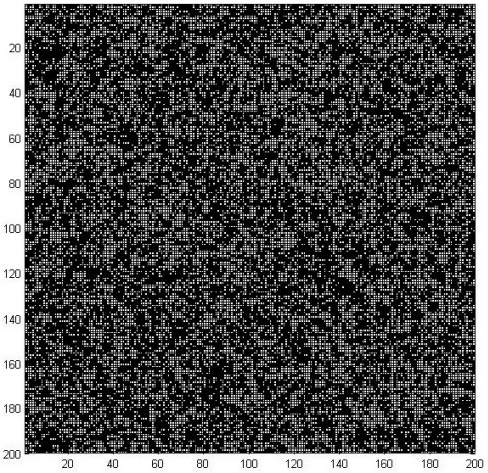
n=500,$v_m$=35



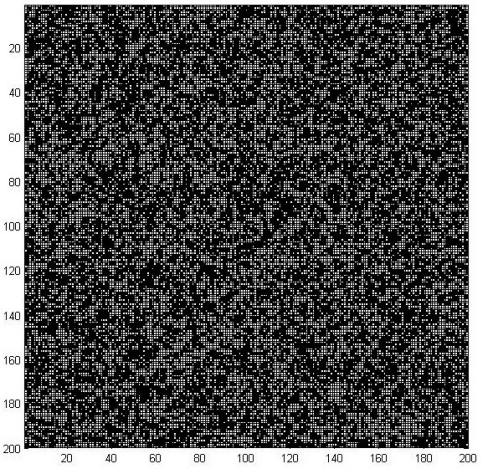

n=1000,$v_m$=50



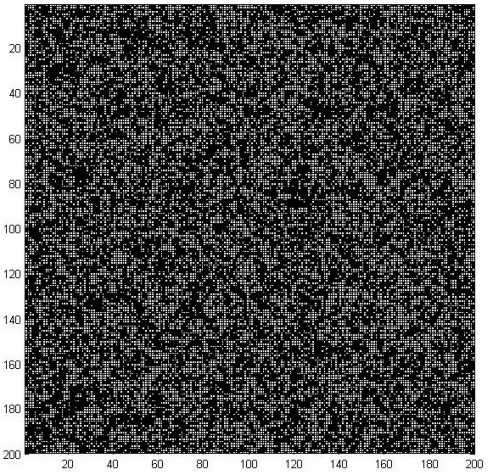

n=500,$v_m$=46



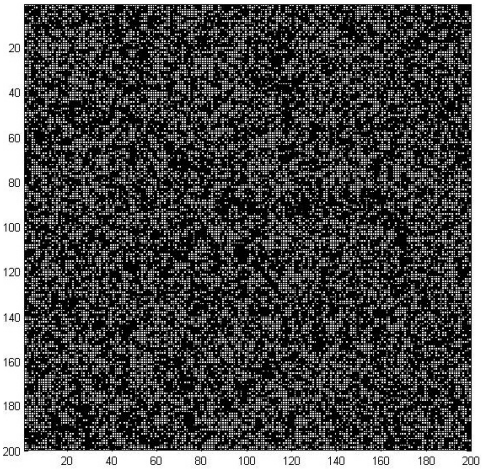

n=1000,$v_m$=65




n=500,$v_m$=50



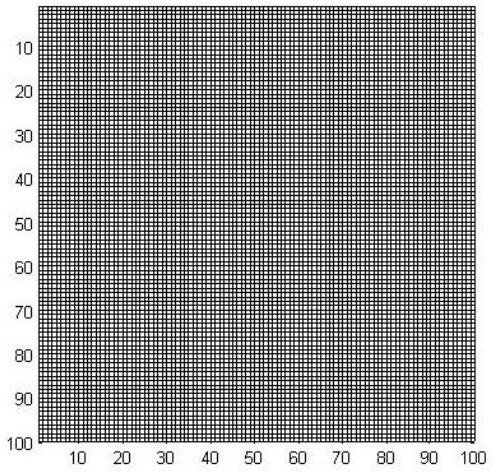

n=1000,$v_m$=70



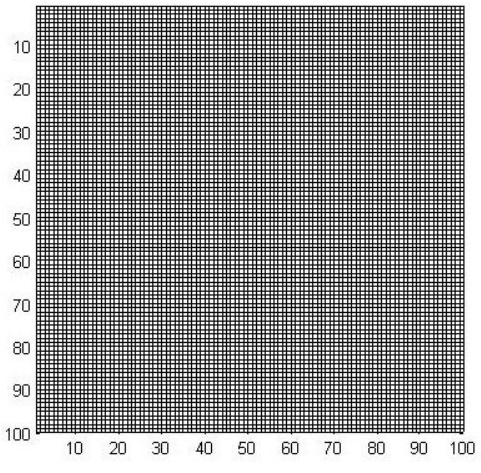
Now the following pictures will show how the bound $v_4$ matters, here again $n = 500, 1000$, and the initial conditions will be chosen that all the particles will be initially in the middle of a half arc (say the white one), the velocity will be given randomly so the role of $v_4$ in this situation will very clear.

TABLE 6.2: the role of bound $v_4$

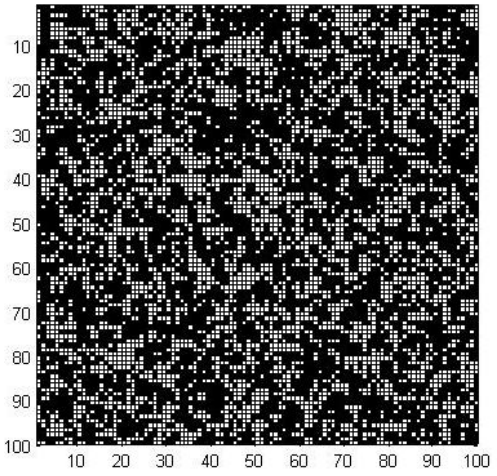
n=500,$v_m$=39



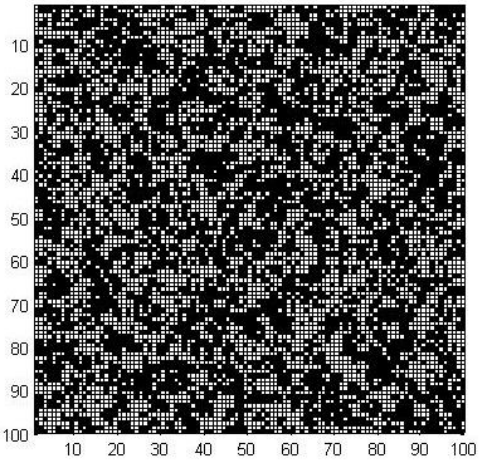

n=1000,$v_m$=57




n=500,$v_m$=40



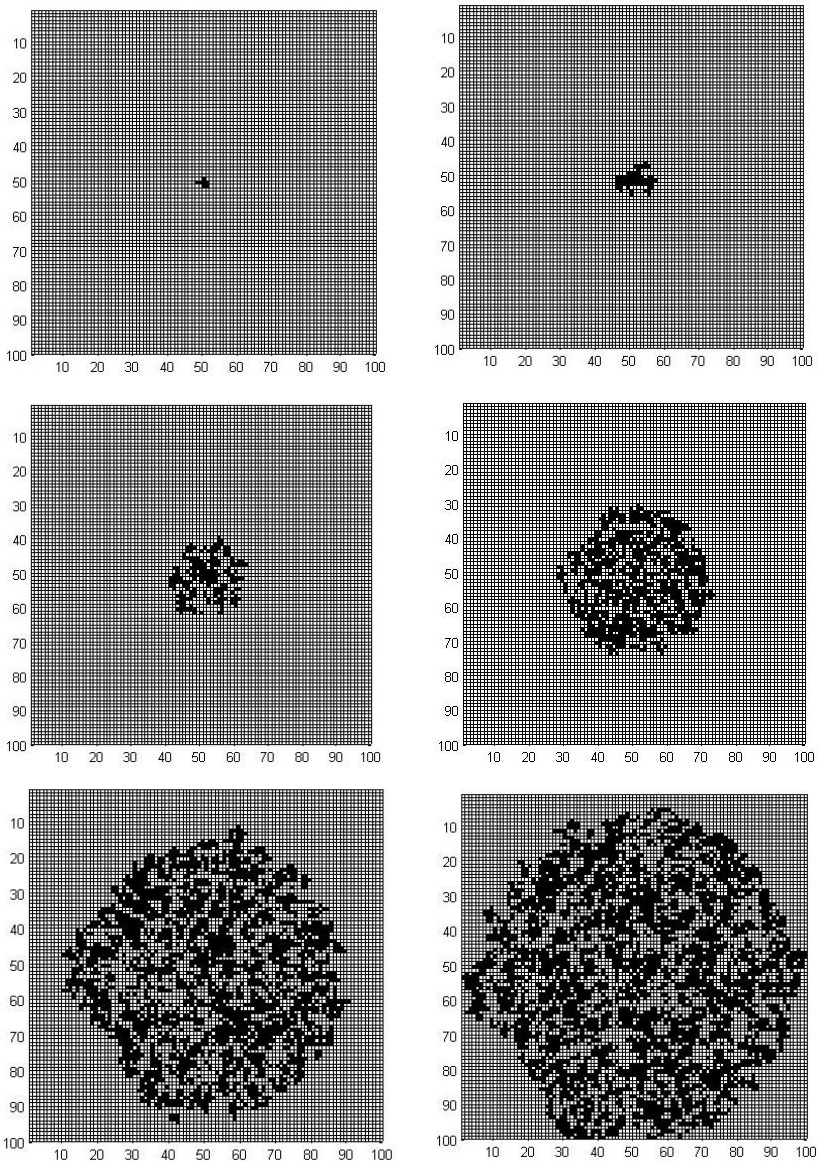

n=1000,$v_m$=58



### 6.3.2 diffusion phenomena

For $v_3 < v_m < v_4 + 1$, and the initial condition is such that some certain particles in the middle of the square are fixed to be black, the other particles are all initially white, then without the fixed black particles the system will be always white there is no case that any particles can change the color, with the fixed black points for $v_m < v_3$ since the cluster will have stable boundaries, the cluster will not actually grow up without any bound, but for $v_4 + 1 > v_m > v_3$ we can observe an interesting diffusion phenomena, the speed of the diffusion will depend on $v_m$ in the sense that the larger the $v_m$ the faster the diffusion process. The results will be shown in next pictures. Here $n = 1000, v_m = 57$

TABLE 6.3: diffusion phenomena



### 6.3.3 non zero external force

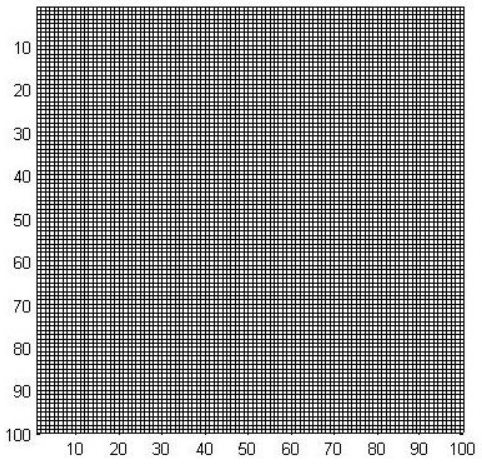
Here are the results for the case when $b$ is not zero, actually $b = 1, n = 1000$ throughout this subsection, in this case we can get some bounds for $v_m$ for which the system will behave different across the bounds, now it is also possible that the equal in color of the 4-neighbours (2 black, 2 white) can trap the particle in the middle. So there are corresponding $v_2, v_3, v_4$.

For $n = 1000$ we have $v_2 = 30, v_3 = 50, v_4 = 63$.

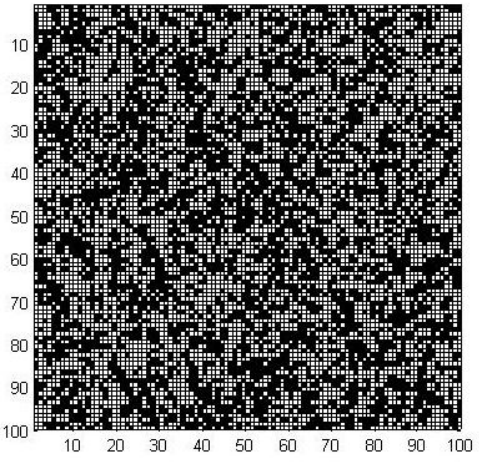
If $v_m > 63$ we would observe that even $b = 1$ the system will still not be ordered, however if $v_m < 50$ then as the case for 0 external force, there will be stable clusters with stable boundaries, but the existence of the external force will make the clusters grow or shrink with preference of one certain color, so in the end there will be one color completely filling the whole space.

TABLE 6.4: non zero external force simulation

n=1000,$v_m$=49,b=1




n=1000,$v_m$=60,b=1




## 6.4 Possible generations and alternatives of this model

1.Instead of the two color system, we can also generalize the system to any number of colors, the force will be defined similarly but in this case the external force will have difference direction in correspond with the preference of the color.

2.We can actually manipulate the form of the interactions between particles in the sense

we can change the integer field induced by the neighbour particles as we wish.

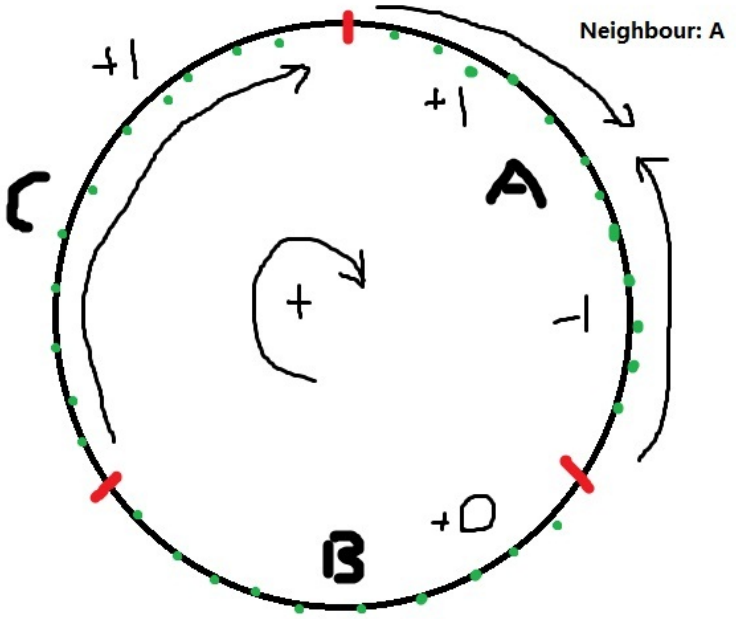
3.Instead of the circle (cyclic group with one generator) we can choose some more complicated algebraic structures for example instead of the 1-cyclic group we had before we can have a 2-cyclic group, can be seen as the 2d version of discrete circle we had before, and the interactions between the particles can be then defined in many ways and I will not go to much in detail about this here.

### 6.4.1 A dynamical model for rock-paper-scissors game

An interesting generalization of the dynamical model will give us a similar result of nowadays popular model in statistical physics and biological physics–the rock-paper-scissors game. A recent research can be found in [Reichenbach,Mobilia,Frey].

The dynamics can be seen in the following picture:

FIGURE 6.2: rock-paper-scissors game dynamics visualized




Intuitively we have 3 different groups of points, named as A,B,C, the dynamics is such that the particle of type A wants to turn its neighbour of type C to A and want to keep

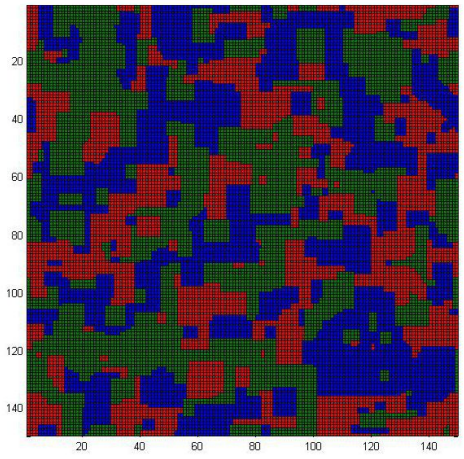
its neighbour staying in A, and the same thing for B to A and for C to B, in an easier illustration:
$CA \to A$, $AB \to B$, $BC \to C$,

Then again the arguments before for the existence of stability of clusters and boundaries will be still valid here, and in this situation if the system has stable cluster with stable boundary the result will be such that the 3 colors divide the square which is not so interesting, and for system without stable cluster we can observe some certain patterns in the system like the shock wave in rock-paper-scissors game in [33].

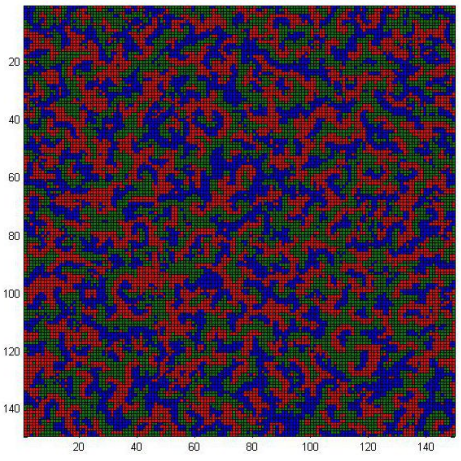
### 6.4.2 simulation results for rock-paper-scissors model

TABLE 6.5: RPS game simulation

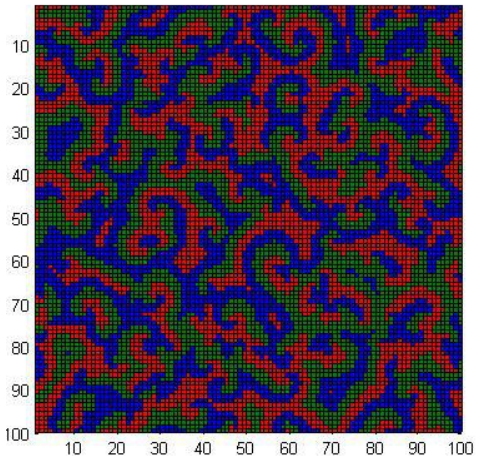
stable case




stable cluster with unstable boundary




unstable case (shock wave)



# Bibliography


[1] Andrzej Lasota and Michael C. Mackey. *Probabilistic Properties of Deterministic Systems*. Cambridge University Press, 1985. ISBN 9780511897474. URL http://dx.doi.org/10.1017/CBO9780511897474. Cambridge Books Online.

[2] Michael C. Mackey. *Time's Arrow: The Origins of Thermodynamic Behavior*. Dover Publications; Dover Ed edition (November 17, 2003), 2003. ISBN 9780486432434.

[3] N. Dunford and J.T. Schwartz. *Linear Operators: General theory*. Linear Operators. Interscience Publishers, 1958. URL http://books.google.de/books?id=i2EPAQAAMAAJ.

[4] D. Ruelle. *Statistical mechanics: rigorous results*. Mathematical physics monograph series. W. A. Benjamin, 1969. URL http://books.google.de/books?id=_LbvAAAAMAAJ.

[5] Ya. G. Sinai I. P. Cornfeld, S. V. Fomin. *Ergodic Theory*. Grundlehren der mathematischen Wissenschaften. Springer New York, 1982.

[6] Boris Hasselblatt Anatole Katok. *Introduction to the Modern Theory of Dynamical Systems*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1996.

[7] P.I. Brown and J.R. Brown. *Ergodic theory and topological dynamics*. Pure and Applied Mathematics. Elsevier Science, 1976. ISBN 9780080873862. URL http://books.google.de/books?id=3nyfnKsWoRkC.

[8] Ya G Sinai. Dynamical systems with elastic reflections. *RUSS MATH SURV*, 25(2) 137–189.

[9] C.J. Thompson. *Mathematical statistical mechanics*. Series of books in applied mathematics. Macmillan, 1971.

[10] H. Furstenberg, Y. Katznelson, and D. Ornstein. The ergodic theoretical proof of szemerédi's theorem. *Bulletin (New Series) of the American Mathematical Society*, 7(527-552)(3):527–552, 11 1982.

[11] Terence Tao Ben Green. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167(481-547).

[12] M. Einsiedler and T. Ward. *Ergodic Theory: with a view towards Number Theory*. Graduate Texts in Mathematics. Springer, 2010. ISBN 9780857290212.

[13] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Pub Co; 2 edition, 1960. ISBN 9780828400237.

[14] J.Clerk Maxwell. On the stresses in rarified gases arising from inequalities of temperature. *Phil. Trans. R. Soc. Lond*, 170(199–245):231–256, 1879.

[15] G.M. Kremer. *An Introduction to the Boltzmann Equation and Transport Processes in Gases*. Interaction of Mechanics and Mathematics. Springer, 2010. ISBN 9783642116964.

[16] L. Boltzmann. *Vorlesungen über Gastheorie: Th. Theorie des Gase mit einatomigen Molekülen, deren dimensionen gegen die Mittlere weglänge verschwinden*. Vorlesungen über Gastheorie. J. A. Barth, 1896.

[17] C.A. Truesdell and R.G. Muncaster. *Fundamentals of MaxwellÆs kinetic theory of a simple monatomic gas: Treated as a branch of rational mechanics*. Pure and Applied Mathematics. Elsevier Science, 1980. ISBN 9780080873992.

[18] S. Chapman and T.G. Cowling. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction, and Diffusion in Gases*. Cambridge University Press, 1958.

[19] Paula Ida Reichert. Can a parabolic evolution of the entropy of the universe provide the foundation for the second law of thermodynamics? Master's thesis, LMU, 2012.

[20] P. Schram. *Kinetic Theory of Gases and Plasmas*. Fundamental Theories of Physics. Springer Netherlands, 1991. ISBN 9780792313922.

[21] J. Bricmont. Science of chaos or chaos in science? *Annals of the New York Academy of Sciences*, 775(131-175):131–175, 1995.

[22] D. Dürr and S. Teufel. *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory*. Springer, 2009. ISBN 9783540893448.

[23] O.Reynolds. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Phil. Trans. R. Soc. Lond*, 186(123-164):123–164, 1895.

[24] Alan Hájek. Interpretations of probability, 2002. URL http://plato.stanford.edu/entries/probability-interpret/.

[25] G.G. Emch and C. Liu. *The Logic of Thermostatistical Physics*. Physics and astronomy online library. Springer, 2002. ISBN 9783540413790.

[26] *Mathematical Interpretation of Formal Systems. [Af] Th. Skolem (o.fl.a.). (Symposion .... Organized by the "Wiskundig Genootschap" .... at Amsterdam on September 9 and 10, 1954*. 1955.

[27] P.G. Hinman. *Fundamentals of Mathematical Logic*. Ak Peters Series. Taylor & Francis, 2005. ISBN 9781568812625.

[28] A.A. Karatsuba, S.M. Voronin, and N. Koblitz. *The Riemann Zeta-Function*. De Gruyter Expositions in Mathematics. De Gruyter, 1992. ISBN 9783110886146.

[29] J.L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart*. Classics in Mathematics. Springer Berlin Heidelberg, 2001. ISBN 9783540412069.

[30] M. Li and P.M.B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Springer, 2009. ISBN 9780387498201.

[31] Richard .L. Epstein and Walter Alexandr Carnielli. *Computability: Computable Functions Logic and the Foundations of Math*. Cole Mathematics Series. Taylor & Francis, 1989. ISBN 9780534103569.

[32] R. Cori and D.L. Lascar. *Mathematical Logic: A Course with Exercises*. Number Bd. 2. 2001.

[33] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Self-organization of mobile populations in cyclic competition. *Journal of Theoretical Biology*, (2):368–383, 2008. URL http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-18396-9.