Università degli studi di Padova

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Laurea Magistrale in Fisica

Are living systems poised at a critical state?

Relatore:
Prof. Amos maritan
Correlatori:
Dott. Samir Suweis
Dott. Jordi Hidalgo
Controrelatore:
Prof. Enzo Orlandini

Laureando:
Marco Faggian

Anno Accademico 2014/2015

Contents

To my family, with love.

Chapter 1

Preface: The Physics of Living Systems

Understanding a living system is a formidable many-body problem. One has an interacting system, made up of different elements (species, neurons, genes...) with imperfectly known interactions, operating in a wide range of spatial and temporal scales. Although such great complexity, the large volume of recently available data made evident that, despite their diversity, living systems exhibit characteristic patterns and regularities, that can be found repeated from microscopic to global scales. For this reason, a considerable part of the physics community has shifted their attention to the study of animate matter with the aim of unraveling the interactions among their constituent elements and the mechanisms responsible of such emergent behavior [57].

Very often, the observed phenomenology in biological systems exhibits similar patterns of other "well-studied" physical systems, and scientist have tried to understand the connection behind, using a combination of approaches inspired by, for instance, statistical mechanics [53], stochastic processes [4] and network analysis [3, 5].

A key attribute of a living organism is its ability to transform its environment, and, in spite of the variety of climate and habitat conditions, we can find robust trends. One of such characteristic patterns corresponds to power-law distributed statistics [17]. Power laws are ubiquitous in living systems, although they span 21 orders of magnitude in mass, with examples ranging from gene and protein networks [59] to cell growth at individual level, and from bacterial clustering [58] to flock dynamics [58] at community level.

One of the lessons taken from statistical mechanics is that power-laws are the fingerprint of criticality at the edge of a continuous phase transition [62], which occurs by a fine tuning of external conditions such as temperature, pressure, chemical potential, magnetic field, etc.. For this reason, the observed phenomenology has lead to conjecture that living systems might be poised at the vicinity of a critical point [2], and it has been argued that this provides robustness and adaptability to best cope with ever-changing environments [2, 52].

Living systems also exhibit long-range correlations between their constituents. For example, in bird flocks, its can be observed that every individual movement is correlated with the others, and that every change has a direct effect on the entire community [60]. Anyone observing the evolution of a flock of starlings [60], a school of fish [17] or a swarm of insects [61] can be fascinated by the synchronization and cohesion among thousands of individuals moving 'in unison'.

In some cases, such behavior can be observed when individuals try to escape from a particular danger, as for instance from a predator, and the group appears to move as macroscopic organism with life and intelligence, as an emergent collective entity. This is the idea

of the "more is different" [57]: a collection of individuals generates a collective behaviour that can be very different from that of the individual itself, and it can be rationalized as the result of simple (individual or local) rules. In the case of bird flocks, every individual interacts with its neighbors trying to align its direction of motion without precise knowledge of the global structure of the entire group [58]. This tendency is propagated by generating a collective coordination without centralized control, i.e. without a leader. Decisions within group such as a turn or a landing are made through an amplification of a local one. Such decentralized coordination is known as self-organization and is widespread in the animal kingdom [58]. Birds, fish and insects give rise to extraordinary developments mainly by their ability to move freely in three dimensions, but there are also numerous examples of terrestrial animals, which move in two dimensions, such as zebras and buffalo. Self-organized collective behavior is also widespread in humans, as occur with the phenomenon of overcrowding – in which a large numbers of people must be evacuated from a confined region (e.g., after a concert or a football game)–, pedestrian traffic and automotive, or the applause at the end of a show.

Physicists, economists, and biologists have developed individual-based models, i.e. that specify the behaviour of each individual with a set of dynamical rules. Formally, each individual is represented by an equation describing the dynamics on the basis of interactions with peers according to some simple rules. From their analytical and numerical study we have learn that from local and simple rules one can get a complex collective behaviour, thus providing a qualitative evidence of one of the fundamental concepts of the theory of self-organization. Quantitative aspects depend on the details of the model, by the way in which the rules are implemented and the value of the various parameters (e.g. magnitude of the interaction zone, importance of the neighbors according to their distance, etc.).

One important case of study is the brain: analyzing the properties properties of neuronal networks, many power law distributions emerge [2]. In particular, the activity of the neuronal networks is examined through what physicists call "neuronal avalanches", which represent the neuronal activity in a time interval and give informations about the quantity of neurons activated in that time period.

Neuroscience is a scientific branch that studies the brain and also tries to understand the behavior of neurons.

Reproducing neuronal activity is a hard task, both from the point of view of modeling and simulation: the brain is a perfect example of a many-body system with interactions between the constituents. Simulating such type of system is complicated and sometimes give an analytical prediction is a great challenge.

In this master's thesis we analyze the collective behavior in neuronal systems: neural activity exhibits power law statistics and, recently, physicists hypothesized that the brain may be posed at a critical state. The goal of this thesis is to provide an overview of recent works on neuronal networks and investigate if they could really be poised at criticality, possibly giving some contributions to this scope. In particular, we study a neuronal model similar to the one described by Johnatan Touboul and Alain Destexhe [4], with which they demonstrate that a neuronal network can exhibit power law statistics even if it is not poised at a critical state. The aim of this work is to search an answer to the following question: "are living systems poised at a critical state?".

First, we will study a neuronal network where neurons are completely uncorrelated, secondly we will study a network with correlated neurons. We will see how these correlations are important in a network to obtain a power law statistics.

This model allows us to study the statistical properties of the neuronal activity: in particular its duration and the quantity of neurons activated in a determined time interval. Finally, we search some clues of criticality following the method of "counting states" described by Thierry Mora et Al. [53].

Chapter 2

Criticality

What is criticality in physics? What do we mean when we refer to critical phenomena? It is useful to introduce formally the concept of criticality –to which we referred in the first chapter– in the framework of statistical mechanics.

2.1 Criticality in Physics

We now see a brief presentation of some models that can undergo a continuous phase transitions (i.e. that have a critical point) and that will be useful latter in our study of biological systems.

2.1.1 The Unidimensional Ising Model

We now consider the one-dimensional Ising model [62] (at first neighbours), consisting of a linear pattern of N nodes, and then with N spins. Suppose that the chain is "closed", i.e. is constituted by $N+1$ spin with $s_{N+1} = s_1$; the Hamiltonian is:

$$H(\{s\}) = -\sum_{i=1}^{N} J_{s_i s_{i+1}} - \sum_{i=1}^{N} h_i s_i \tag{2.1}$$

Let's just consider the case of uniform chain and field ($J_i = J$ and $h_i = h$) and rewrite Eq.(2.1) as:

$$H(\{s\}) = -\sum_{i=1}^{N} \left[J_i s_i s_{i+1} + \frac{1}{2} h (s_i + s_{i+1}) \right] \tag{2.2}$$

With the method of "transfer matrix", we observe that the partition function is

$$Z(T, h, N) = \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} e^{-\beta H(\{s\})} \tag{2.3}$$

It can be written, using the condition of closing of the chain, such as:

$$Z(T, h, N) = \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} \mathcal{T}_{s_1 s_2} \mathcal{T}_{s_2 s_3} \dots \mathcal{T}_{s_N s_1} \tag{2.4}$$

where we have defined: $a \equiv \beta J$ and $b \equiv \beta h$ and

$$\mathcal{T}_{s_i s_{i+1}} = \exp(a s_i s_{i+1} + \frac{b}{2}(s_i + s_{i+1})) \quad i = 1, \dots, N \tag{2.5}$$

The numbers $T_{s_{i+1}}$ (for $s_i = \pm 1$, $s_{i+1} = \pm 1$) can be seen as elements of the 2×2 matrix:

$$\mathcal{T} = \begin{pmatrix} \mathcal{T}_{+1+1} & \mathcal{T}_{+1-1} \\ \mathcal{T}_{-1+1} & \mathcal{T}_{-1-1} \end{pmatrix} = \begin{pmatrix} e^{a+b} & e^{-a} \\ e^{-a} & e^{a-b} \end{pmatrix} \tag{2.6}$$

The expression Eq.(2.4) is the sum of the diagonal elements of the matrix $\mathcal{T}^N$:

$$Z = Tr\left[\mathcal{T}^N\right] \tag{2.7}$$

We observe that $\mathcal{T}$ is real and symmetric. The eigenvalues of this matrix are

$$\lambda_\pm = e^a \left[\cosh b \pm \sqrt{\sinh^2 b + e^{-4a}}\right] \tag{2.8}$$

It is evident that

$$\lambda_+ > 0 \quad \lambda_- \geq 0 \tag{2.9}$$

and then in general it is proved that

$$\lambda_+ > \lambda_- \tag{2.10}$$

The partition function is then:

$$Z = \lambda_+^N + \lambda_-^N \tag{2.11}$$

We limit ourselves to the region $T > 0$. The average value of a spin can be written as:

$$\langle s_i \rangle = \frac{1}{Z} \sum_{i=1}^{N} s_i e^{-\beta H(\{s\})} = \frac{1}{\beta} \frac{1}{Z} \frac{\partial Z}{\partial h_i} \tag{2.12}$$

The first important result is the *mean magnetization* of the system:

$$\langle s \rangle_i = \frac{\sinh b + \frac{\sinh b \cosh b}{\sqrt{\sinh^2 b + e^{-4a}}}}{\cosh b \pm \sqrt{\sinh^2 b + e^{-4a}}} \tag{2.13}$$

For a uniform chain $M = N \langle s \rangle_1$. Consider then the "magnetization", that is the "magnetic moment to spin", in the thermodynamic limit $N \to \infty$ we obtain

$$m(T, h) = \lim_{N(\Omega) \to \infty} \frac{M}{N} = \lim_{N(\Omega) \to \infty} \langle s_1 \rangle \tag{2.14}$$

Consider an Ising model with any size, with uniform interaction (nearest neighbour), and with non-uniform field. The Hamiltonian becomes:

$$H(\{s\}) = -\frac{J}{2} \sum_{i \neq j = 1}^{N} s_i s_j - \sum_{i=1}^{N} h_i s_i \tag{2.15}$$

We call *correlation functions* the mean values $\langle s_i s_j \rangle$. The functions $\langle s_i s_j \rangle$ can be expressed by a formula analogous to

$$\langle s_i s_j \rangle = \frac{1}{\beta^2} \frac{1}{Z} \frac{\partial^2 Z}{\partial h_i \partial h_j} \tag{2.16}$$

For $J = 0$, the correlation functions assume a particular form

$$\langle s_i s_j \rangle = \langle s_i \rangle \langle s_j \rangle \quad i \neq j, \quad J = 0 \tag{2.17}$$

For $J = 0$ the spins are "independent" or "uncorrelated". Consider then correlation functions "related" G_{ij}, defined as:

$$G_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle \tag{2.18}$$

The G_{ij} are related to the *local susceptibility* χ_{ij}:

$$\chi_{ij} = \frac{\partial}{\partial h_j} \langle s_i \rangle = \frac{1}{\beta} \frac{\partial}{\partial h_j} \left(\frac{1}{Z} \frac{\partial Z}{\partial h_i} \right) \tag{2.19}$$

χ_{ij} represents the variation of the mean value $\langle s_i \rangle$ as the result of a variation of the field at the point j. In the absence of interaction $\chi_{ij} = 0$. The relationship between G_{ij} and χ_{ij} is

$$G_{ij} = k_B T \chi_{ij} \tag{2.20}$$

Experimentally has been found that, for a system of ferromagnetic uniform, there are a "critical temperature" T_c, and a length, $\xi = \xi(T)$, said *correlation length*, such that in the *critical region*:

$$G(r) \simeq r^{-\alpha} e^{-r/\xi} \tag{2.21}$$

where α is a positive number. ξ characterizes the rapidity with which G_{rij} tends to zero, i.e. the rapidity with which $\langle s_i s_j \rangle$ tends to $\langle s_i \rangle \langle s_j \rangle$.
Of great importance is the dependence of ξ by T: experimentally, in the critical point, $(T, h) = (T_c, 0)$,

$$G(r) \propto r^{-\alpha}. \tag{2.22}$$

Then, for $T = T_c$ we have $\xi = \infty$.
In the same model, the finite size *free energy* is defined as:

$$F_\Omega(T, \{h_i\}, \{J_{ij}\}, \dots) = -k_B T \ln(Tr e^{-\beta H_\Omega}) \tag{2.23}$$

In the thermodynamic limit the free energy per site is:

$$f_b(T, \{h_i\}, \{J_{ij}\}, \dots) = \lim_{N(\Omega) \to \infty} \frac{1}{N(\Omega)} F_\Omega(T, \{h_i\}, \{J_{ij}\}, \dots) \tag{2.24}$$

In the one-dimensional case, the free energy in the thermodynamic limit is

$$f_b(T, \{h_i\}, \{J_{ij}\}, \dots) = -k_B T \lim_{N \to \infty} \frac{1}{N} \ln(\lambda_+^N + \lambda_-^N) \simeq -k_B T \ln \lambda_+ \tag{2.25}$$

The result for the correlation lenght is

$$\xi^{-1} = \ln \frac{\lambda_+}{\lambda_-} \tag{2.26}$$

What is essential is to compute the eigenvalues and then to obtain the precise value of these quantities. For example,

$$\xi_{H=0}^{-1} = \frac{1}{\ln(\coth(a))} \tag{2.27}$$

In the limit $T \to 0$ ($a \to \infty$), as

$$\coth(a) \simeq 1 + 2e^{-2a} \tag{2.28}$$

which provides

$$\xi \simeq e^{J/k_B T} \quad T \to 0 \tag{2.29}$$

for $T \to 0$ the correlation length diverges exponentially (essential singularity in zero).

2.2 Why is criticality so special?

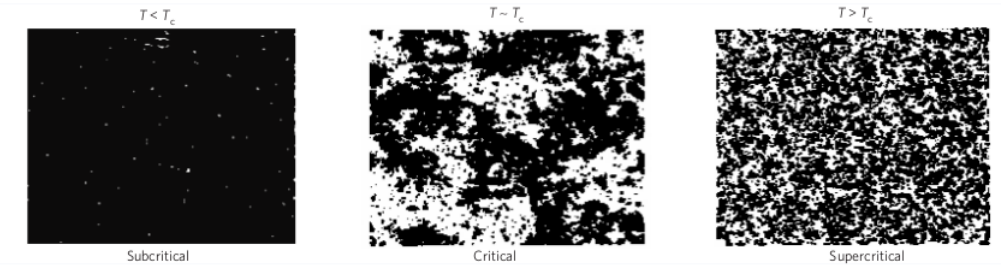


Figure 2.1: Three snapshots of the spin configurations at one moment in time for three temperatures (subcritical, critical and supercritical) from numerical simulations of the two-dimensional Ising model.[51]

In Fig.2.1 are represented three snapshots of the spin configurations at different temperatures.
Let us recall what is special about the critical state considering the case of an iron magnet. When it is heated, the magnetization decreases until it reaches zero beyond a critical temperature T_c . Individual spin orientations are, at high temperatures, changing continuously in small groups. As a consequence, the mean magnetization, expressing the collective behaviour, vanishes. At low temperature the system will be very ordered, exhibiting large domains of equally oriented spins, a state with negligible variability over time. In between these two homogeneous states, at the critical temperature T_c , the system exhibits very different properties in both time and space. The temporal fluctuations of the magnetization are known to be scale invariant. Similarly, the spatial distribution of correlated spins shows long-range (power-law) correlations. It is only close enough to T_c that large correlated structures (up to the size of the system) emerge, even though interactions are with nearest elements.
In addition, the largest fluctuations in the magnetization are observed at T_c . At this point the system is at the highest susceptibility, a single spin perturbation has a small but finite chance to start an avalanche that reshapes the entire system state, something unthinkable in a non-critical state.

Many of these dynamical properties, once properly translated and adapted to neuronal terms, exhibit striking analogies to brain dynamics. For instance, neuromodulators, which are known to alter brain states acting globally over non-specific targets, could be thought of as control parameters, as is temperature in this case. We will refer to this issue later.

2.2.1 Continuous Phase Transitions and Critical Exponents

The physics of continuous phase transitions is studied under the framework of statistical mechanics. A phase transition involves a variation of the physical properties of the system such as its density, electrical conductivity, magnetization, the crystalline structure and so on. In the presence of a continuous phase transition there will be a singularity of the free energy and / or its derivatives.

In a nutshell, to characterize a phase transition we refer to a physical quantity that represents, with some exceptions, the major qualitative difference between one phase and another. This concept, introduced by Landau in 1937, is called *order parameter*. Each phase transition is characterized by an order parameter. The order parameter is a physical variable that gives a measure of the level of order in a system. In most cases it describes many of the rearrangements of the structure that take place due to the phase transition, and it allows

to distinguish between the two phases involved in the transition. It can be a size scalar, a vector, or in some cases, a magnitude tensor.

Critical exponents are defined at the critical point of the continuous phase transition. We should first define the reduced temperature t of the system in such a way as to have a dimensionless quantity that takes the form

$$t = \frac{T - T_c}{T_c} \tag{2.30}$$

In particular, t is a measure of the deviation from the critical temperature T_c. If $t \to 0$ it means that the system is at a temperature very close to the critical temperature. Let suppose that we have a thermodynamic function $F(t)$ continuous and positive for sufficiently small values of t. To ensure the positivity of the reduced temperature t take the amount in module. It is assumed that the following limit exists and is finite:

$$\lambda = \lim_{t \to 0} \frac{\ln F(t)}{\ln |t|}. \tag{2.31}$$

λ is called the *critical exponent* of a thermodynamic function and features critical transitions. In any critical transition each thermodynamic function is characterized by a well-determined critical value of the exponent. It is easy to see that

$$F(t) \sim |t|^{\lambda} \tag{2.32}$$

In correspondence of the critical point where we have that $t \to 0$ the thermodynamic function presents alternatively two behaviors: if $\lambda < 0$ it diverges, if $\lambda > 0$ it resets. The divergence is both right and left over at the critical point. The critical value of the exponent then expresses the degree of departure or the instance seeking annulment of the thermodynamic function. Examples of divergent behavior at the critical point can be the magnetic susceptibility at constant temperature χ_T, the specific heat calculated to constant volume for a fluid (C_V) and constant magnetic field for a magnetic system (C_H).

The critical exponents that characterize the behavior of the different functions in correspondence of the thermodynamic critical point are typically different and are indicated with the greek letters of the alphabet α, β, γ, δ and so on (see Table 2.1).

One of the most important thermodynamic functions is the susceptibility, which diverges

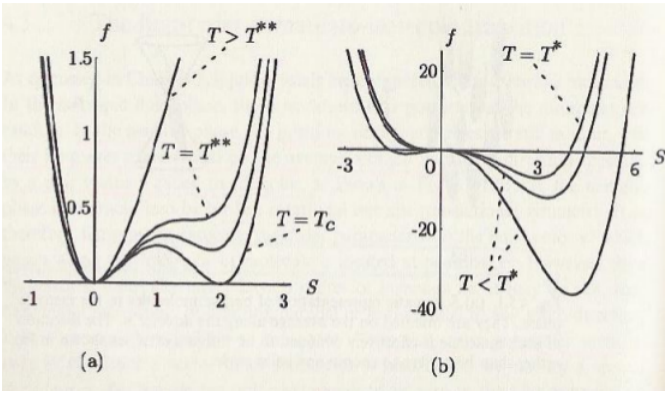


Figure 2.2: Free energy density f as function of order parameter for different T for isotropic-nematic transition. The transition is first order. Note the limits of metastability for supercooling (T^*) and superheating (T^{**}). [8]

Quantity	Behavior	Region	Exponent
C_h	$\|T-T_c\|^{-\alpha}$	$h=0$	$\alpha = 0(discont.)$
χ_r	$\|T-T_c\|^{-\gamma}$	$h=0$	$\gamma = 1$
m	$\|T-T_c\|^{\beta}$	$h=0$	$\beta = 1/2$
	$h^{1/\delta}$	$T=T_c$	$\delta = 3$
ξ	$\|T-T_c\|^{-\nu}$	$h=0$	$\nu = 1/2$

Table 2.1: Examples of critical exponents for some quantities for a thermodynamic system in the Landau Theory (Fig.2.2).

at criticality for large systems. Let us apply a very small magnetic field to the system in the critical point. A very small magnetic field is not able to magnetize a large coherent cluster, but with these fractal clusters the picture changes. It affects easily the smallest size clusters, since they have a nearly paramagnetic behaviour. But this change affects also the next-scale clusters, and the perturbation climbs the ladder until the whole system changes radically. Thus, critical systems are very sensitive to small perturbations.

In the case of Landau Theory (Fig.2.2), there exists an analytical function that describes the behavior of the free energy density of a system posed near a critical point.

This function, called $\mathcal{L}$, is of the following form:

$$\mathcal{L}(\eta) \sim a_0 + a_1\eta + a_2\eta^2 + a_3\eta^3 + a_4\eta^4 + \cdots + O(\eta^n) \tag{2.33}$$

where $\mathcal{L}$ is a function of η, a order parameter of the system. In the case of the Ising model, η is the magnetization and it is scalar. The critical exponents in this model are reported in Table 2.1.

2.2.2 Universality

Given the definition of critical exponents we have to justify why they are interesting from a physical point of view. The critical temperature depends strongly on the type of interactions that describe the system in which a phase transition occurs. Instead critical exponents are universal, they only depend of a few basic parameters. For systems that can be studied using models with short range interactions, they depend on the dimensionality of the space and on the symmetry of the order parameter.
This means that the critical exponents of one-dimensional systems are different from those of two-dimensional and three-dimensional systems, and so on. For example the two-dimensional Ising model was solved analytically by Onsager for $H = 0$, $\beta = 1/8$, while for the model of three-dimensional Ising numerical calculations provide exactly $\beta = 0,33$. But they have the same values of critical exponents at fixed dimensionality of the systems regardless of the nature of the type of substance being studied. In addition, for a fixed dimensionality, different systems in which the Hamiltonian has the same symmetry with respect the order parameter have the same critical exponents. The symmetry of the order parameter depends of what type of magnitude are we representing, i.e. if it is a scalar, a vector or a tensor.

We can introduce the concept of *universality class*, which is a group of systems that present the same critical exponents. Universality is one of the most important predictions of the renormalization group, thanks to which the thermodynamic properties of a system near the critical phase transition expressed by the critical exponents (determinable exactly by the renormalization group) depend only on a small number of characteristics. The most important consequence of the universality is the one for which, in order to determine the

critical exponents of a class of systems (phenomena), we can just consider the simplest system. This much simpler system does not necessarily have a physical feedback.

2.3 Homogeneous functions of one or more variables

It is seen that when the system approaches a critical point ξ the distance along which the fluctuations are related becomes comparable to the size of the entire system and the microscopic effects as space lattice or the size of the fluid particles become irrelevant. We can expect that in the vicinity of a critical point, if we move a bit from it (for example by changing slightly the temperature) the free energy does not change in its form but only its scale. The idea of scaling is the basis of calculation of all the critical exponents. To understand the scaling is appropriate to introduce the concept of homogeneous function.

Definition 1. Taken $\mathcal{A}$ an open set of $\mathbb{R}^n$ and α a real number, a function $f : \mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be homogeneous of α degree in $\mathcal{A}$ if

$$f(tx) = t^\alpha f(x) \qquad \forall x \in \mathcal{A}, \forall t > 0 \tag{2.34}$$

More generally, if

$$f(tx) = g(t)f(x) \qquad \forall x \in \mathcal{A}, \forall t > 0 \tag{2.35}$$

where $g(t)$ is a general power-law function of degree α.

An example of homogeneous function with $\alpha = 2$ in $\mathbb{R}$ is the parable $f(x) = Bx^2$:

$$f(tx) = Bt^2x^2$$

An homogeneous functions with $\alpha = 0$ in $\mathcal{A} = \{(x, y,) \in \mathbb{R}^2 : y \neq 0\}$ is

$$f(x, y) = \frac{x}{y}$$

It can be shown that the function $g(t)$ must be only of the following form:

$$g(t) = t^\alpha$$

where α is said "*degree of homogeneity*".

Proof. Suppose to do a change of scale before in μ and after in t, starting from the equation

$$f(tx) = g(t)f(x) \tag{2.36}$$

it follows

$$f[t(\mu x)] = g(t)f(\mu x) = g(t)g(\mu)f(x) \tag{2.37}$$

but also

$$f[(t\mu)x)] = g(t\mu)f(x) \tag{2.38}$$

Suppose now that $g(t)$ is differentiable. Differentiating with respect to the μ is obtained

$$tg'(t\mu) = g(t)g'(\mu) \tag{2.39}$$

We put $\mu = 1$ and define $\alpha \equiv g(\mu = 1)$. So we have

$$tg'(t) = g(t)\alpha \tag{2.40}$$

Integrating the differential equation and using the properties of the same equation, we obtain

$$g(t) = t^\alpha \tag{2.41}$$

□

The extension to the case of n variables is natural and leads to the definition of an homogeneous function with n variables:

$$f(t\bar{x}) = g(t)f(\bar{x}). \tag{2.42}$$

2.4 High variability: Heavy-Tailed or Scaling Distributions

Focusing on non-negative random variables X, let $F(X) = P[X \leq x], x \geq 0$, denote the *cumulative distribution function* (CDF) of X and $\bar{F}(x) = 1 - F(x)$ the complementary CDF (CCDF).
A typical feature of commonly-used distribution functions is that their (right) tails decrease exponentially fast, implying that all moments, including exponentials moments, exist and are finite. To describe in a mathematically convenient way high variability phenomena, we introduce the class of *subexponential distribution functions*. Following Goldie and Kluppelberg (1998), we call F (or X) *subexponential* if

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{e^{-\varepsilon x}} \to \infty \quad \forall \varepsilon > 0$$

that is, the (right) tail of a subexponential distribution decays more slowly than any exponential, implying that all exponential moments of a subexponential are infinite.
Well-known examples of subexponential distributions include the Lognormal, Weibull, Pareto of the first or second kind, and stable laws, while Gaussian, exponential and Gamma are examples that are not in the class of subexponential.
To distinguish between subexponential distributions we consider the subclass of these kind of distributions consisting of the *heavy-tailed* or *scaling* distributions. A subexponential distribution function $F(x)$ or random variable X is called *heavy-tailed* or *scaling* if for some $\alpha \in (0, 2)$

$$P[X > x] \sim cx^{-\alpha} \quad as \quad x \to \infty$$

where $c \in \mathbb{R}_+$. The parameter α is called *tail index*, for $\alpha \in [1, 2)$ F has infinite variance but finite mean; for $\alpha \in (0, 1)$ F has not only infinite variance, but also infinite mean. In general, all moments of order $\beta > \alpha$ are infinite:

$$< x^\beta > \simeq c \int_x^{+\infty} y^{-\alpha+\beta}\, dy = \begin{cases} +\infty & if \quad \beta \geq \alpha \\ \\ < +\infty & if \quad \beta < \alpha \end{cases}$$

Heavy-tailed distributions are called scaling distributions because the sole response to conditioning is a change in scale; that is, if X is heavy-tailed with index α and $x > \omega$, the conditional distribution of X given that $X > \omega$ satisfies

$$P[X > x | X > \omega] = \frac{P[X > x]}{P[X > \omega]} \simeq c_1 x^{-\alpha}$$

which is identical to the unconditional distribution $P[X > x]$, except for a change in scale. For example, the non-heavy tailed exponential distribution gives

$$P(X > x | X > \omega) = e^{-\lambda(x-\omega)}$$

that is, the conditional distribution is also identical to the unconditional, except for a change of location rather than scale.

An important feature of scaling distributions that distinguishes them from their commonly-considered non heavy-tailed counterparts concerns their mean residual lifetime defined as $E(X-x|X>x)$. For an exponential distribution with constant parameter λ (given $x = e^{\lambda\omega}$)

$$E(X-x|X>x) = \int_x^{+\infty} e^{-\lambda(x-\omega)} = \frac{1}{\lambda}$$

In stark constrast for scaling distributions the mean residual lifetime is linearly increasing:

$$E(X-x|X>x) = \int_x^{+\infty} (X-x)x^{-\alpha} \simeq cx$$

2.5 Invariance properties

Scaling distributions enjoy a number of invariance properties that characterize them. Following the presentation of Mandelbrot (1997), we are going to show that these types of distributions are invariant under transformations as aggregation, mixture, maximization and marginalization and discuss some practical implications of these properties.

2.5.1 Aggregation

The classical central limit theorem is often cited as the reason for which Gaussian distributions occur in nature.

Theorem 2.5.1. *Suppose that $(X_n : n \geq 1)$ is a sequence of IID random variables with distribution function F, where F has finite mean m and finite variance σ^2. Let $S_n = X_1 + \ldots X_n, n \geq 1$ denote the n^{th} partial sum. Then, as $n \to \infty$,*

$$n^{-1/2}(S_n - mn) \sim \sigma N(0,1)$$

where $N(0,1)$ is the standard Gaussian distribution having mean 0 and unitary variance.

For a somewhat less-konwn version of the Central Limit Theorem (CLT), we recall that a random variable U is said to have a *stable law (with index $\alpha \in (0,2]$)* if for any $n \geq 2$, there is a real number d_n such that

$$U_1 + U_2 + \ldots U_n = n^{1/\alpha}U + d_n$$

where $U_1, U_2, \ldots U_n$ are independent copies of U. Stable laws on the real line can be represented as a four-parameter family $S_\alpha(\sigma, \beta, \mu)$, with the index $\alpha \in (0,2]$, the *scale parameter* $\sigma > 0$, the *skewness parameter* $\beta \in [-1,+1]$ and the *location (shift) parameter* $\mu \in (-\infty, +\infty)$.
When $\alpha \in (1,2)$ the shift parameter is the mean, but for $\alpha \leq 1$, the mean is infinite. While for $\alpha < 2$ all stable laws are heavy-tailed, the case $\alpha = 2$ is special and represents a familiar, not heavy-tailed distribution: the Gaussian (normal) distribution, $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$.

Theorem 2.5.2. *Suppose that $(X_n : n \geq 1)$ is a sequence of non-negative, IID random variables with scaling distribution function F with $\alpha \in (1,2)$ (implying finite mean m but infinite variance σ^2). Let $S_n = X_1 + \ldots X_n, n \geq 1$ denote the n^{th} partial sum. Then, as $n \to \infty$,*

$$n^{-1/\alpha}(S_n - mn) \Rightarrow S_\alpha(1, \beta, 0)$$

Both Gaussian and scaling distributions are invariant under aggregation, as proved by these theorems. In general terms, classical and non-classical CLTs state that the stable distributions are the only fixed points of the renormalization group transformations (i.e. aggregation) and that Gaussian distributions are, as a matter of fact, a very special case.

2.5.2 Maximizing Choices

Consider n independent random variables denoted $X_1, X_2, \ldots X_n$ and assume that their distribution functions are scaling distributions with the same index α, but possibly with different scale coefficients $c_i > 0$; that is,

$$P(X_i > x) \simeq c_i x^{-\alpha} \quad for \quad i \in [1, n]$$

For $k \in [1, n]$, define the random variables M_k to be the k-th successive maximum given by

$$M_k = max(X_1, X_2, \ldots, X_k)$$

Using that $P(M_k \leq x) = \prod_{i \in [1,k]} P(X_i \leq x)$, it is easy to show that for large x,

$$P[M_k > x] \simeq c_{M_k} x^{-\alpha}$$

where $c_{M_k} = \sum_{i \in [1,k]} c_i$. Thus, the k-th successive maxima of scaling distributions are also scaling, with the same tail index α, but different scale coefficients than the individual X_i's. Scaling distributions are the only type that are invariant under the transformation of maximization.

2.5.3 Weighted Mixtures

Consider n independent random variables denoted $X_1, X_2, \ldots X_n$ with scaling distribution functions F_i, all with the same tail index α but different scale coefficients $c_i > 0$. Consider now the *weighted mixture* W_n of the X_i's and denote with p_i the probability that $W_n = X_i$. It is easy to show that

$$P[W_n > x] = \sum p_i P[X_i > x] \simeq c_{W_n} x^{-\alpha}$$

where $c_{W_n} = \sum p_i c_i$ is the *weighted average* of the separate scale coefficients c_i. The distribution of the weighted mixture is also scaling. The converse (i.e. W_n is scaling only if the X_i's are scaling) holds only to a first approximation. In the limit $n \to \infty$ the invariant distributions for W are of the form $P[W > x] = cx^{-\alpha}$, $x \geq 0$, which are improper distribution functions because they yield an infinite total probability. However, for all practical purposes, the X_i's are typically restricted by some relation of the form $0 < a < x$ which results in perfectly well-defined (conditional) distribution functions of the scaling type.

Marginalization

Recall that stable distributions are trivially scaling. The multivariate stable distributions can be characterized as being those for which every linear combination of the coordinates has a (scalar) stable distribution. We call this transformation marginalization.

2.6 Widom scaling theory

It was seen previously that when there is a phase transition in a system, a part of free energy behaves so that some response functions of the system have a singularity often in the form of divergences. If we assume that the non-analytical free energy does not change shape but only scale, then it is possible to generalize the inequalities between critical exponents found in thermodynamics with real relationships that agree with experiments.
Letting $t = (T - T_c)/T_c$ and $h = (H - H_c)/(k_B T)$, suppose to write the free energy density in two parts

$$f(T, H) = f_r(T, H) + f_s(t, h) \tag{2.43}$$

where $f_r(T, H)$ is the regular part (does not change significantly as we approach the point critic) and $f_s(t, h)$ is the singular part that contains the singular behavior of the system the vicinity of the critical point, that is, for $(t, h) \sim (0, 0)$. The hypothesis of static scaling asserts that the singular part of the free energy $f_s(t, h)$ is a homogeneous generalized function:

$$f_s(\lambda^{p_1} t, \lambda^{p_2} h) = \lambda f_s(t, h) \quad \forall \lambda \in \mathbb{R} \tag{2.44}$$

The assumption that the Gibbs free energy (for example) is a homogeneous generalized function implies that all other thermodynamic potentials are homogeneous generalized functions. With an appropriate choice of parameter λ, one of the arguments of the function can be removed. For example in our case if we choose

$$\lambda = h^{-1/p_2} \tag{2.45}$$

we would have

$$f_s(t, h) = h^{1/p2} f_s\left(\frac{t}{h^{p_1/p_2}}, 1\right) \tag{2.46}$$

In literature the relationship $p2/p1 \equiv \Delta$ is said *gap exponent*.

Now let's see what are the implications of Eq. (2.44) on the performance of the thermodynamic quantities near the critical point and, ultimately, on relations between critical exponents. Let's start with the magnetization. Since the magnetization is the first derivative of the free energy with respect to the magnetic field, we derive both sides of Eq. (2.44) with respect to h (would than H but the factor β is eliminated being common to the two members). This provides

$$\lambda^{p_2} \frac{\partial f_s(\lambda^{p_1} t, \lambda^{p_2} h)}{\partial h} = \lambda \frac{\partial f_s(t, h)}{\partial h} \tag{2.47}$$

and so

$$\lambda^{p_2} M(\lambda^{p_1} t, \lambda^{p_2} h) = \lambda M(t, h) \tag{2.48}$$

There are two critical exponents β and δ associated with the behavior of the magnetization in the vicinity the critical point. Consider first the exponent β that corresponds to put $h = 0$ and $t \to 0^-$. For $h = 0$ Eq. (2.48) becomes

$$M(t, 0) = \lambda^{p_2 - 1} M(\lambda^{p_1} t, 0) \tag{2.49}$$

Since it is true for every value of the parameter λ, we eliminate the dependence on the second function putting $\lambda = t^{-1/p_1}$. Then

$$M(t, 0) = (-t)^{(1-p2)/p1} M(-1, 0). \tag{2.50}$$

Moreover, as $t \to 0^-$ we have $M \sim (-t)^\beta$, we get

$$\beta = \frac{1 - p_2}{p_1} \tag{2.51}$$

The exponent δ is had for $T = Tc$ $(t = 0)$ and $h \to 0$. We have

$$M(0, h) = h^{(1-p2)/p2} M(0, 1). \tag{2.52}$$

Moreover $h \to 0$ $M \sim h^{1/\delta}$ and comparing the two reports is obtained

$$\delta = \frac{p_2}{1 - p_2} \tag{2.53}$$

We also see that the gap is the ratio exponent

$$\Delta = \frac{p_2}{p_1} = \beta\delta \tag{2.54}$$

2.7 Linear Relation between Energy and Entropy

The fundamental variables of thermodynamics are energy, temperature, and entropy. For the states taken on by a network of neurons, energy and temperature are meaningless, so it is difficult to see how we can construct a thermodynamics for these systems. But in statistical mechanics, all thermodynamic quantities are derivable from the Boltzmann distribution, the probability that the system will be found in any particular state.
We start by recalling that, for a system in thermal equilibrium at temperature T , the probability of finding the system in state s is given by

$$P_s = \frac{1}{Z} e^{-E_s/k_B T} \tag{2.55}$$

where E_s is the energy of the state, and Boltzmann's constant k_B converts between conventional units of temperature and energy. The partition function Z serves to normalize the distribution, which requires

$$Z = \sum_s e^{-E_s/k_B T} \tag{2.56}$$

but in fact this normalization constant encodes many physical quantities.

The logarithm of the partition function is proportional to the free energy of the system, the derivative of the free energy with respect to the volume occupied by the system is the pressure, the derivative with respect to the strength of an applied magnetic field is the magnetization, and so on.
Thermodynamics doesn't make reference to all these details. Which aspects of the underlying microscopic rules actually matter for predicting the free energy and its derivatives? We can write the sum over all states as a sum first over states that have the same energy, and then a sum over energies. We do this by introducing an integral over a delta function into the sum:

$$Z = \sum_s e^{-E_s/k_B T} = Z = \sum_s e^{-E_s/k_B T} \left[\int dE \delta(E - E_s) \right] = \tag{2.57}$$

$$= \int dE e^{-E_s/k_B T} \left[\sum_s \delta(E - E_s) \right] \tag{2.58}$$

where $\delta(x)$ is the Dirac delta function.
We see that the way in which the energy depends on each state appears only in the brackets, a function $n(E)$ that counts how many states have a particular energy.
Instead of counting the number of states that have energy E, we can count the number of states with energy less than E:

$$\mathcal{N}(E) = \sum_s \theta(E - E_s) \tag{2.59}$$

where the step function is the Heaviside's Theta and is defined as

$$\theta(x) = \begin{cases} 1 & x > 0 \\ \\ 0 & x < 0 \end{cases} \tag{2.60}$$

But the step function is the integral of the delta function, which means that we can integrate by parts in Eq.(2.58) to give

$$Z = \frac{1}{k_B T} \int dE e^{-E_s/k_B T} \mathcal{N}(E) \tag{2.61}$$

If we think about N variables, each of which can take on only two states, the total number of states is 2^N. More generally, we expect that the number of possible states in a system with N variables is exponentially large, so it is natural to think not about the number of states $\mathcal{N}(E)$ but about its logarithm,

$$S(E) = \ln \mathcal{N}(E) \tag{2.62}$$

which is called the *entropy*.
As a technical aside, we can define the entropy either in terms of the number of states with energy close to E, what we have called $n(E)$ in the main text, or we can use the number of states with energy less than E, what we have called $\mathcal{N}(E)$. In the limit that the number of degrees of freedom in the system become large, there is no difference in the resulting estimate of the entropy per degree of freedom, because the number of states is growing exponentially fast with the energy, so that the vast majority of states with energy less than E also have energy very close to E.
Using these results, we can define the partition function as a function of the energy and the entropy:

$$Z = \frac{1}{k_B T} \int dE e^{[-E/k_B T + S(E)]} \tag{2.63}$$

It is then natural to think about the energy per particle $\varepsilon = E/N$, and the entropy per particle, $S(E)/N = s(E)$. In the limit of large N , we expect $s(E)$ to become a smooth function. We can also write Z as

$$Z = \frac{N}{k_B T} \int d\varepsilon e^{-N f(\varepsilon)/k_B T} \tag{2.64}$$

where $f(\varepsilon) = -k_B T s(\varepsilon)$.
We note that $f(\varepsilon)$ is the difference between energy and entropy, scaled by the temperature, and is called free energy.
Whenever we have an integral of the form in Eq.(2.64), at large N we expect that it will be dominated by values of close to the minimum of $f(\varepsilon)$. This minimum ε_* is the solution to the equation

$$\frac{df(\varepsilon}{d\varepsilon} = 0 \Rightarrow \frac{1}{k_B T} = \frac{ds(\varepsilon)}{d\varepsilon} \tag{2.65}$$

which we can also think of as defining the temperature. Notice that T being positive requires that the system have $ds(\varepsilon)/d\varepsilon > 0$, which means there are more states with higher energies.
We can expand $f(\varepsilon)$ in the neighborhood of ε_*

$$f(\varepsilon) = f(\varepsilon_*) - \frac{k_B T}{2} \left(\frac{d^2 s(\varepsilon)}{d\varepsilon^2} \right)_{\varepsilon_*} (\varepsilon - \varepsilon_*)^2 \tag{2.66}$$

With the interpretation of ε_* as the mean energy per particle, we can calculate how this energy changes when we change the temperature, and we find

$$\frac{d\varepsilon_*}{dT} = \frac{1}{k_B T^2} \left[- \left(\frac{d^2 s(\varepsilon)}{d\varepsilon^2} \right)_{\varepsilon_*} \right]^{-1} \tag{2.67}$$

The change in energy with temperature is called the heat capacity C, and when we normalize per particle it is referred to as the specific heat.
Combining equations,

$$\langle (\delta\varepsilon)^2 \rangle = k_B T^2 \frac{C}{N} \tag{2.68}$$

Our discussion assumes that the second derivative of the entropy with respect to the energy is not zero. If we take all our results at face value, then when $d^2s/d^2\varepsilon \to 0$, the specific heat will become infinite, as will the variance of the energy per particle. This is a critical point.

There is much more to be said about the analysis of critical points using the entropy versus energy. But our concern here is how these ideas connect to systems that are not in thermal equilibrium, so that temperature and energy are not relevant concepts.
Rather than trying to compute the partition function, we can ask, for any distribution, how the normalization condition is satisfied. We still imagine that there are states s, built of of N different variables, as with the patterns of spiking and silence in a network of neurons. Each state s has a probability P_s , and we must have

$$1 = \sum_s P_s \tag{2.69}$$

Using a thermodynamic approach, we can define

$$E_s = -\ln\left(\frac{P_s}{P_0}\right) \tag{2.70}$$

where P_0 is the probability of the most likely state.
In the case of interest to us here, we can look at the states taken on by groups of N neurons, and we can vary N over some range. The function $\mathcal{N}(E)$, and hence the entropy $S(E)$, is a property of a single system with a particular value of N, and to remind us of this fact we can write $S_N(E)$.
It is an experimental question what happens as N become large. But, in many of the examples we understand—from statistical physics, from information theory, and indeed from more general examples in probability theory—we find that there is a well defined limiting behavior at large N , which means that there is a function

$$s(\varepsilon) = \lim_{N\to\infty} \frac{1}{N} S_N(E = N\varepsilon) \tag{2.71}$$

We can assign an "energy" to every state of the system, which is just the negative log probability. It is convenient to normalize so that the most likely state has zero energy. The "effective temperature" of the system is $k_BT = 1$.
We can count the number of states below a given energy, and the log of this number is an entropy. If there are N elements (e.g., neurons) in our system, it is natural to ask about the entropy per element as a function of the energy per element. If this function has a smooth limit as N becomes large, $s(\varepsilon)$, then we can define a thermodynamics for the system.
The large variance in log probability is mathematically equivalent to a diverging specific heat in the thermodynamic case. This is a signature of a critical point. In equilibrium systems with interactions that extend only over short distances, correlations typically extend over some longer but finite distance ξ; at the critical point this correlation diverges, so that there is no characteristic length scale—all scales between the size of the constituent particles and the size of the system as a whole are relevant [40].
Not only does the specific heat diverge at the critical point, but so does the susceptibility to external fields. All of these diverging quantities have a power–law dependence on the difference between the actual temperature and the critical temperature, and the exponents of these power–laws are quantitatively universal: many different systems, with different microscopic constituents, exhibit precisely the same exponents, and in a certain precise sense these exponents give a complete description of the system in the neighborhood of the critical point [41][42].

In the study of complex, non–equilibrium systems, scale invariance and power–law behaviors often are taken as signs of criticality, but seldom is it possible to exhibit these behaviors over the wide range of scales that are the standard in studies of equilibrium critical phenomena, so one must be cautious. Deterministic dynamical systems also exhibit critical phenomena, often called bifurcations, where the system's behavior changes qualitatively in response to an infinitesimal change in parameters [43]. These phenomena are easiest to understand when the number of degrees of freedom N is small, but then the sharp bifurcations are rounded if there is noise in the system; the example of equilibrium statistical mechanics shows how noisy dynamical systems can recover sharp transitions in the limit of large N.

Chapter 3

Are Living Systems Poised at Criticality?

Many of most fascinating phenomena in living organisms emerge from interactions among many elements: amino acids determine the structure of a single protein, genes determine the fate of a cell, neurons are involved in shaping our thoughts and memories... In the past few years, new, larger scale experiments have made possible to develop accurate models of biological systems directly from real data.
The stationary states of biological systems have a subtle structure, neither "frozen" into a well ordered crystal, nor chaotic and disordered like a gas. These states are far from equilibrium, maintained by a constant flow of energy and material through the system. For example, it was suggested that the brain is in a self-organized critical state, at the edge between being a nearly dead (frozen state) and being fully epileptic (fully disordered state) [63].
In the last years physicists have developed different models to understand the collective behavior of living systems, with their power-law statistics and long-range correlations. One of the most relevant framework in our context is the theory of *self-organized criticality.* [**?**]

The theory of self-organized criticality [64] has its origin in models of inanimate matter (sandpiles, earthquakes, etc.), but its has been extended and adapted to encompass biological systems through the analysis of simple toy models. As an example, simple models for the evolution of interacting species can self-organize to a critical state in which periods of quiescence are interrupted by "avalanches" of mutations, which reminds us of the idea of punctuated equilibria in evolution [55].

However, several questions still remain open:

- What is the advantage of a living system of being poised at a critical state?
- From a biological point of view, what is the meaning of "being critical"?
- For many systems we know that if we would just pick model parameters at random, we will not find anything that reproduces biological functions. Thus, biological systems operate in special regions of the parameter space. What does define the special region of parameters? And how can we identify the significative order parameters of the system?

All these questions have not a definitive answer today.

In many cases one can identify criticality in purely thermodynamic measurements, as a singularity in the heat capacity as a function of temperature, or through the behavior of the correlation function of fluctuations in some local variables, such as the magnetization in a magnet. The difficulty in biological systems is that they are not really equilibrium statistical mechanics problems, so there is no guarantee that we can find some relevant macroscopic variables, and certainly it is not clear what is the meaning of "temperature", while the correlation function can be defined also for a system out of equilibrium.

Now we review some models that support the hypothesis that living systems are poised at a critical state and other models that are against it. We start now from the arguments in favor of criticality.

3.1 Pro Criticality

3.1.1 Flocks of Birds

Groups of animals such as schooling fish, swarming insects or flocking birds move with fascinating coordination. Rather than being dictated by a leader or responding to a common stimulus, the collective patterns of flock dynamics tend to be self-organized, and arise from local interactions between individuals, which propagate information through the whole group [58].

Flocks, schools and swarms are also highly responsive and cohesive in the face of predatory threat. This balance between order and chaos directly points to the idea of criticality. A recent analysis has framed this idea in precise mathematical terms, culminating in the first empirical evidence that flock behavior may indeed be critical in the sense of statistical physics [61].

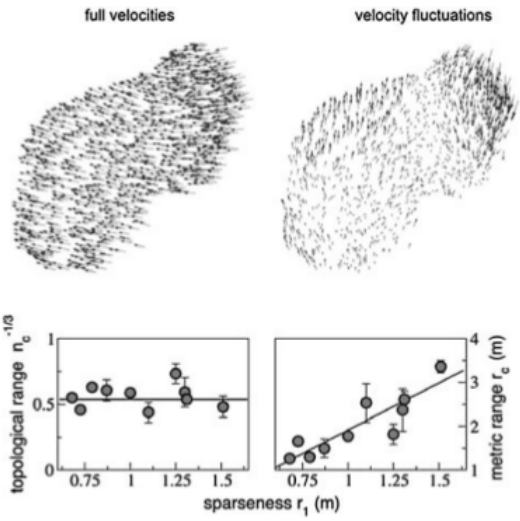


Figure 3.1: Two dimensional projection of a typical 3D reconstruction of the positions and velocities of every bird in a flock of 1,246 starlings. Left: the absolute velocities $\bar{v}_i$ show a high degree of order in bird orientation. Right: the velocity fluctuations, $\bar{u}_i = \bar{v}_i - \frac{1}{N}\sum_{i=1}^{N} \bar{v}_i$, are long-ranged, and form only two coherent domains of opposite directions.[2]

Birds interact with their neighbors according to their topological distance (measured in

units of average bird separation), rather than to their metric distance (measured in units of length). The distribution of neighbors around an average bird is not uniform: birds tend to have closer neighbors on their sides than behind or in front of them. There are biological reasons for this. Birds have lateral vision, and can monitor their lateral neighbors with better accuracy. In addition, keeping a larger distance with frontal neighbors may be a good strategy for avoiding collisions. The main assumption is that this heterogeneity is a result of interactions between individuals, and can be used to estimate the range of these interactions, defined as the distance at which the neighborhood of an average bird becomes uniform.
How does global order emerge across the whole flock from local interactions? Clearly, if each bird perfectly mimics its neighbors, then a preferred orientation will propagate without errors through the flock, which will align along that direction. Alignment with neighbors is not perfect, and noise could impede the emergence of global order.

Interestingly, a phase transition occurs in simple models of flock dynamics as in the case of a uniform, fully connected Ising model [60].

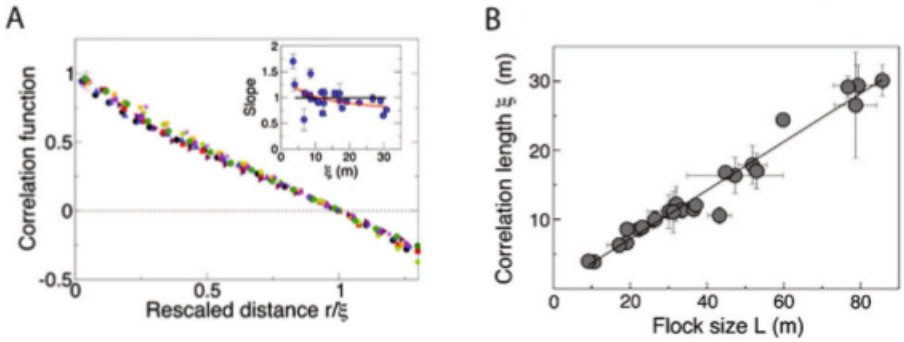


Figure 3.2: Velocity fluctuations are scale free. A. The correlation length ξ scales linearly with the system's size L, indicating that no other scale than L is present in the system. B. Correlation function C versus rescaled distance r/ξ . ξ is defined as the radius for which $C = 0$. The slope at $r = \xi$ (inset) seems to depend only weakly upon ξ . This suggests that coherence can in principle be preserved over extremely long ranges.[2]

For this system, near a critical point, correlation functions are given by a universal function [60],

$$C(r) = \frac{1}{r^{\gamma}} f(r/\xi) \tag{3.1}$$

where ξ is the correlation length which diverges as the critical point is approached.
Replacing $\xi = \alpha L$ into Eq.(3.1) and taking $L \to \infty$ yields a power law decay for the correlation function, $C(r) = r^{-\gamma}$, characteristic of a critical point.
Flocks form a cohesive mass and long range order may appear as a natural consequence of this cohesion. To see why long range correlations are in fact surprising, let us contrast flocks with a well understood case: a solid. As a solid moves, the positions of its atoms evolve in a highly coordinated and correlated manner. However, the thermal fluctuations of these positions are only weakly correlated with each other across long distances. By contrast, in flocks not only birds do fly in the same general direction, but their small variations from that direction are strongly correlated over the entire extent of the flock. [60]

3.1.2 Emergence of Criticality in Living Systems

Evidence has been mounting that biological systems might operate at the borderline between order and disorder, i.e., near a critical point.

Jorge Hidalgo et Al. [52], rationalize this apparently ubiquitous criticality in terms of adaptive and evolutionary functional advantages. They provide an analytical framework, which demonstrates that the optimal response to broadly different changing environments occurs in systems organizing spontaneously—through adaptation or evolution—to the vicinity of a critical point.

Their goal of [52] was to exploit general ideas from statistical mechanics and information theory to construct a quantitative framework showing that self-tuning to criticality is a convenient strategy adopted by living systems to effectively cope with the intrinsically complex external world in an efficient manner, thereby providing an excellent compromise between accuracy and flexibility. The main result is that criticality is an evolutionary/adaptive sta-

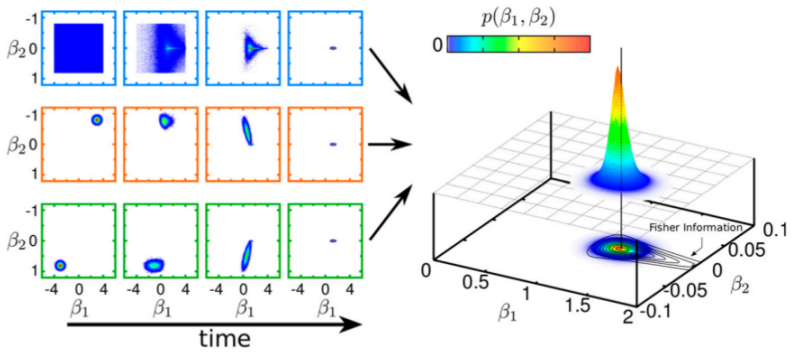


Figure 3.3: Coevolutionary model leads self-consistently to criticality: A community of M living systems (or agents) evolves according to a genetic algorithm dynamics. Each agent i ($i = 1, \ldots, M$) is characterized by a two-parameter (β_1^i, β_2^1) internal state distribution $P_{int}(s|\beta_1^i, \beta_2^1)$, and the rest of the community acts as the external environment it has to cope with, i.e., the agents try to "understand" each other. At each simulation step, two individuals are randomly chosen and their respective relative fitnesses are computed in terms of the KL divergence from each other's internal state probability distribution. One of the two agents is removed from the community with a probability that is smaller for the fitter agent; the winner produces an offspring, which (except for small variations/ mutations) inherits its parameters. (Left) These coevolutionary rules drive the community very close to a unique localized steady state. As shown (Right), this is localized precisely at the critical point, i.e., where the generalized susceptibility or Fisher information of the internal state distribution exhibits a sharp peak (as shown by the contour plots and heat maps). [52]

ble solution reached by living systems in their striving to cope with complex heterogeneous environments or when trying to efficiently coordinate themselves as an ensemble.
The external environment in which living systems operate is highly variable, largely unpredictable, and describable in terms of probability distribution functions. Living systems need to modify their internal state to cope with external conditions, and they do so in a probabilistic manner.

Living systems need to modify their internal state to cope with external conditions, and they do so in a probabilistic manner. In the presence of broadly different ever-changing heterogeneous environments, computational evolutionary and adaptive models vividly illustrate how a collection of living systems eventually clusters near the critical state. A more accurate convergence to criticality is found in a coevolutionary/coadaptive setup in which individuals evolve/adapt to represent with fidelity other agents in the community, thereby creating a

collective "language," which turns out to be critical.
These ideas apply straightforwardly to genetic and neural networks—where they could contribute to a better understanding of why neural activity seems to be tuned to criticality—but have a broader range of implications for general complex adaptive systems.

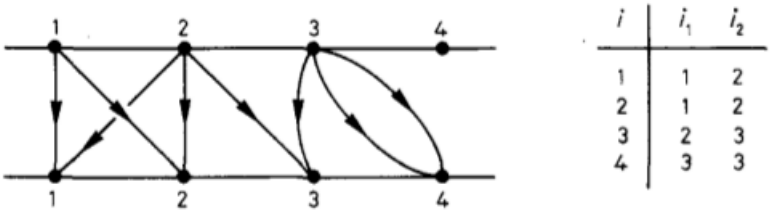
3.1.3 Random Networks of Automata

Now we see a particular type of model for the neuronal networks, analyzed by Deridda and Pomeau [5] in 1985. Kauffman's model is a random complex automata where nodes are randomly assembled. Each node o, receives K inputs from K randomly chosen nodes and the values of σ_i at time $t+1$ is a random Boolean function of the K inputs at time t. Numerical simulations have shown that the behaviour of this model is very different for $K > 2$ and $K \leq 2$. It is the purpose of this work to give a simple annealed approximation which predicts $K = 2$ as the critical value of K.
Networks of Boolean automata to study the behaviour of generic regulatory systems were introduced by Kauffman in 1969. The subsequent studies have revealed surprisingly ordered structures in randomly constructed networks. In particular the most highly organized behaviour (small attractors, smaller number of attractors, stable attractors, etc.) appeared to occur in networks where each node receives inputs from two other nodes.

Let us first briefly describe Kauffman's model. The model depends on a parameter K. The system consists of N spins σ_i , which can take two possible values ($\sigma_i = 0$ or 1). The time evolution of this system is given by N Boolean functions of K variables each

$$\sigma_i^{t+1} = f_i(\sigma_{i_1}^t, \sigma_{i_2}^t \ldots, \sigma_{i_K}^t) \tag{3.2}$$

For each i, the spins $\sigma_{i_1}, \sigma_{i_2} \ldots, \sigma_{i_K}$ are randomly chosen among the N spins. They need not be different (i.e. i_1 can be equal to i_2 for example). So the system is defined once a function f_i and a set $i_1, i_2, \ldots, i_K$ have been chosen for each site i.



i	i_1	i_2
1	1	2
2	1	2
3	2	3
4	3	3

Figure 3.4: An example for $N = 4$ and $K = 2$.[5]

There exist 2^N possible Boolean functions of K variables. In Kauffman's model, each function f_i is randomly chosen among these 2^{2^K} possible functions. The system is random because, for each i, the set $i_1, \ldots, i_K$ is randomly chosen and because the function f_i is random. This randomness is quenched because the functions f, and the sets $i_1, \ldots, i_K$ are quenched (they do not change with time).
There are many questions one can ask about this model:

- What is the length of the limit cycles?
- What is the number of different limit cycles?

If one considers 2 different spin configurations at time $t = 0$, what is the probability that they become the same at time t? If one starts with two randomly chosen configurations $\mathcal{C}_1$

and $\mathcal{C}_2$ of the system at time $t = 0$, what will be the distance between these configurations after time t and in the limit $t \to \infty$?
Let us consider 2 spin configurations $\mathcal{C}_1$ and $\mathcal{C}_2$ which are at distance n. By definition the distance $d(\mathcal{C}_1, \mathcal{C}_2)$ between two spin configurations is n, if the number of spins which are different in the two configurations is n. If one considers two randomly chosen configuration g and such that $d(\mathcal{C}_1, \mathcal{C}_2) = n$ at time $t = 0$, one can calculate the probability $P_1(m, n)$ that the distance $d(\mathcal{C}'_1, \mathcal{C}'_2)$ between their images $\mathcal{C}'_1$ and $\mathcal{C}'_1$ at time $t = 1$ is $d(\mathcal{C}'_1, \mathcal{C}'_2)m$. To calculate $P_1(m, n)$, let us call A the set of spins which are identical in $\mathcal{C}_1$ and $\mathcal{C}_2$ and B the set of spins which are different. Set A contains $N - n$ spins, whereas B contains n spins. Let us call $Q(No)$ the probability that N_0 spins have all their K arrows coming from set A . One has

$$Q(N_0) = \binom{N}{N_0} \left[\left(\frac{N-n}{N} \right)^K \right]^{N_0} \left[1 - \left(\frac{N-n}{N} \right)^K \right]^{N-N_0} \tag{3.3}$$

These N_0 spins will be of course identical in $\mathcal{C}_1$ and $\mathcal{C}_2$. For the remaining $N - N_0$ spins, since at least one of their inputs is different in $\mathcal{C}_1$ and $\mathcal{C}_2$, there is a probability $1/2$ that they are the same in $\mathcal{C}_1$ and $\mathcal{C}_2$ and $1/2$ that they differ. Therefore, the probability $P_1(m, n)$ that $d(\mathcal{C}'_1, \mathcal{C}_2) = m$ is

$$P_1(m, n) = \sum_{N_0=0}^{N} Q(N_0) \left(\frac{1}{2} \right)^{N-N_0} \binom{N - N_0}{N} \tag{3.4}$$

One can easly compute the sum over N_0 and find

$$P_1(m, n) = \frac{N!}{2^N m!(N-m)!} \left[1 + \left(1 - \frac{n}{N} \right)^K \right]^{N-m} \left[1 - \left(1 - \frac{n}{N} \right)^k \right]^m \tag{3.5}$$

If one neglects the correlations and studies the model where at each time the functions f_i and the sets $1_1 \ldots, i_K$ are changed, then $P_t(m, n)$ in this *annealed approximation* is

$$P_t^{annealed}(m, n) = \sum_{q_1=0}^{N} \cdots \sum_{q_{t-1}=0}^{N} P_1(m, q_{t-1}) P_1(q_{t-1}, q_{t-2}) \ldots P_1(q_1, n) \tag{3.6}$$

For large N, can be introduced $x = \frac{n}{N}$ and $y = \frac{m}{N}$ and it can be found that the annealed solution becomes very peaked around a well-defined value of y.

In general y_t is given by

$$y_t = \frac{1 - (1 - y_{t-1})^K}{2} \tag{3.7}$$

where y_1 is given by

$$y_1 = \frac{1 - (1 - x)^k}{2} \tag{3.8}$$

We see that once the distance Nx between-$\mathcal{C}_1$ and $\mathcal{C}_2$ is given, then with probability 1 in the limit $N \to \infty$, the distance between their images at time t is Ny_t. Then the principal problem is reduced to study the Eq.(3.7).
The fixed point can be now found with simple computations:

$$\lim_{t \to \infty} \lim_{N \to \infty} \frac{d(\mathcal{C}_1^t, \mathcal{C}_2^t)}{N} = \begin{cases} 0 & K \leq 2 \\ \\ y^* & K > 2 \end{cases} \tag{3.9}$$

Then this simple approximation gives a critical value $K_c = 2$.

K	y^* fixed point of (9)	Masured distance between 24-th and 25-th iterates for $N = 108$
2	0	0.113 ± 0.097
3	0.38197	0.37 ± 0.078
5	0.48121	0.485 ± 0.047

Figure 3.5: Comparison of the value of y^* with measured distances in simulations of Kauffman's model.[5]

3.1.4 Boolean Networks

The same result was found by Alireza Goudarzi, Christof Teuscher, Natali Gulbahce and Thimo Rohlf [3] in networks learning informations, in their model learning refers to correctly solving the task for the training samples, while generalization refers to correctly solving the task for novel inputs.
They used *adaptation* to refer to the phase where networks have to adapt to ongoing mutations, but have already learned the input-output mapping. They studied adaptive information processing in populations of Boolean networks with an evolving topology.
A Random Boolean Network (RBN) is a discrete dynamical system composed of N automata. Each automaton can have two possible states $\{0, 1\}$, and the dynamics is such that $F : \{0, 1\}^N \to \{0, 1\}^N$, where $f = (f_1, \ldots, f_i, \ldots, f_N)$ and each f_i is represented by a look-up table of K_i inputs randomly chosen from the set of N automata.
They restricted the maximum K_i to 8, the time-evolving rule for a state $\sigma_i^t \in \{0, 1\}$ is the following:

$$\sigma_i^{t+1} = f_i(\sigma_{i_1}^t, \sigma_{i_2}^t, \ldots, \sigma_{i_K}^t) \tag{3.10}$$

They evolved networks by means of a traditional GA to solve three computational tasks of varying difficulty, each of which defined on a 3-bit input: full adder (FA), even-odd (EO) and the cellulare automata rule 85 (R85).
They calculated a weighted sum of the vector of the information content of all possible decomposition models of F and the weight w_m of a model is proportional to its degrees of freedom. The information content of a model os calculated by using $\log f_m = 1 - \frac{H_m - H_f}{H_{ind} - H_f}$, where H_m is the entropy of the model, H_f the entropy of F and H_{ind} is the entropy of the independence model.
It was used only a mutation-based genetic algorithm. For all the experiments it is ran a population of 30 networks with initial connectivity $\langle K_{in} \rangle = 1$ and a mutation rate of 0.8. Each mutation is decomposed into $1 + \alpha$ steps repeated with probability $p(\alpha) = 0.5^{\alpha+1}$, where $\alpha \geq 1$. Each step involves flipping a random location of the look-up table of a random automaton combined with adding or deleting one link.
During each generation, the fitness of each individual is determined by $f = 1 - E_M$, where E_M is the average error over the T random training samples:

$$E_M = \frac{1}{T} \sum_{i \in M} \sum_{j \in O} (a_{ij} - o_{ij})^2 \tag{3.11}$$

where a_{ij} is the value of the output automata j for the input pattern i, and o_{ij} is the correct value of the same bit for the corresponding task. Selection is applied to the population as a deterministic tournament: two individuals are picked randomly from the old population, and their fitness values are compared. The better individual is mutated and inserted into the new population, the worse individual is discarded. They repeated this process until they have 30 new individuals.
In this work they observed (Fig.3.6) a convergence of $\langle K \rangle$ close to the critical value $K_c = 2$

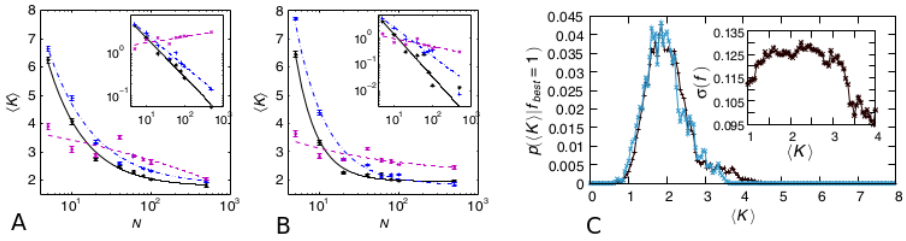


Figure 3.6: Finite size scaling of $\langle K \rangle$ as a function of N for the three tasks, EO (black), FA (blue), R85 (magenta), and the training sample size $T = 4$ A) and $T = 8$ B). Points represent the data of the evolved networks; lines represent the fits. The finite size scaling for $\langle K \rangle$ shows that it scales with a power law as a function of the system size N. The dashed lines represent the power-law fit of the form $a x^b + c$. We favor the data for larger N by weighting the data according to N/N_{max} , where $N_{max} = 500$. The insets show $K_c - c$ as a function of N on a log-log scale. In the Figure C) The conditional probability that evolving populations, where the best mutant reaches maximum fitness (i.e., $f_{best} = f_{max} = 1$), have average connectivity $< K >$ shows a sharp peak near K_c (black curve); the same is found for maximum generalization (light blue). Inset: The diversity of evolving populations, quantified in terms of the standard deviation $\sigma(f)$ of fitness distributions, has a maximum near $K_c = 2$. All data were sampled over the best 22 out of 30 populations for the full-adder task with $T = 4$ and $N = 100$. [3]

for large system size N and training sample sizes larger than or equal to $T = 4$. For the R85 task they don't observe any convergence to $K_c = 2$, for the other tasks is observed a convergence towards K_c with a power law as a function of the system size N. The population dynamics in this model follows Fisher's fundamental theorem of natural selection, which attributes the rate of increase in the mean fitness to the increased fitness variance in the population. The standard deviation has a local maximum near K_c, with a sharp decay toward larger $\langle K \rangle$, indication of maximum diversity near criticality. In this model is found that the population where the best mutant has maximum fitness ($f = 1$) sharply peak ner K_c, as well as populations where the best mutant reaches perfect generalization.

With this work it is showed that learning and generalization are optimized near criticality, furthermore, critical RBN populations exhibit the largest diversity (variance) in fitness values, which supports learning and robustness of solutions under continuous mutations.
Also, examination of the attractors of the final population confirms that the computation happens as partitioning of the state space into disjoint attractors.
During the evolution, the attractor landscape changes so that there are enough attractors to properlyprocess the inputs. (see Appendix A)
Learning of classification tasks and adaptation can drive RBNs to the edge of chaos, where high-diversity populations are maintained and ongoing adaptation and robustness are optimized.

3.1.5 Dynamical Versus Statistical Criticality in Neuronal Networks

Beggs and Plenz [54] were the first to report such power laws in the context of neuronal networks.

Relatively recent work has reported that networks of neurons can produce avalanches of activity whose sizes follow a power law distribution. This suggests that these networks may be operating near a critical point, poised between a phase where activity rapidly dies out and a phase where activity is amplified over time. The hypothesis that the electrical activ-

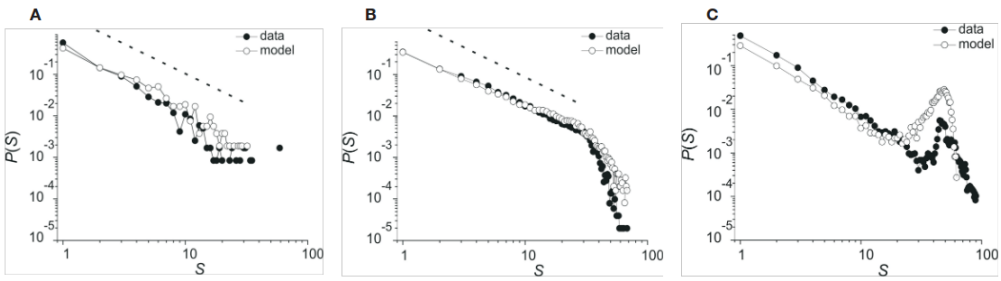


Figure 3.7: Avalanche size distributions in local field potential data collected with a 60-channel microelectrode array from rat cortical slice networks. (A) Subcritical regime; excitatory antagonist applied. (B) Critical regime; normal network. (C) Supercritical regime; inhibitory antagonist applied. [54]

ity of neural networks in the brain is critical is potentially important, as many simulations suggest that information processing functions would be optimized at the critical point. This hypothesis, however, is still controversial.

Consider a control parameter for neuronal excitability, which sets how much a spike in one neuron excites its neighbors. If this parameter is too low, a spike in one neuron may propagate to its direct neighbors, but the associated wave of activity will quickly go extinct. Conversely, if the excitability parameter is too high, the wave will explode through the whole population and cause something reminiscent of an epileptic seizure. To function efficiently, a neuronal population must therefore pose itself near the critical point between these two regimes.

The analogy with sandpiles and earthquakes is straightforward: when a grain falls, it dissipates some its mechanical energy to its neighbors, which may fall in response, provoking an avalanche of events.

The most striking feature of self-organized criticality is the distribution of the avalanche sizes, which typically follows a power law. In their experiment, a 60-channel multielectrode array was used to measure local field potentials (a coarse grained measure of neural activity) in cortical cultures and acute slices.

Activity occurred in avalanches—bursts of activity lasted for tens of milliseconds and were separated by seconds long silent episodes— that propagated across the array (Fig.3.8A). For each event, the total number of electrodes involved was counted as a measure of avalanche size. The distribution of this size s followed a power-law with an exponent close to $-3/2$ (Fig.3.8B). Although that exponent was first speculated to be universal, it was later shown that it depends on the details of the measurement method.

The critical properties of neuronal avalanches can be explained by a simple branching process. Assume that when a neuron fires at time t, each of its neighbors has a certain probability of firing at time $t+1$, such that the average number of neighbors firing at $t+1$ is given by the branching parameter β. That parameter is exactly what we called "excitability" earlier; $\beta < 1$ leads to an exponential decay of the avalanche, $\beta > 1$ to its exponential and unlimited growth, and $\beta = 1$ defines the critical point.

To summarize, in the space of avalanche configurations we have the same signature of criticality that we have seen in the retina, although in different tissues, with different measurement methods, and assuming different models of activity. This emphasizes the potential generality of Zipf's law and criticality for brain function.

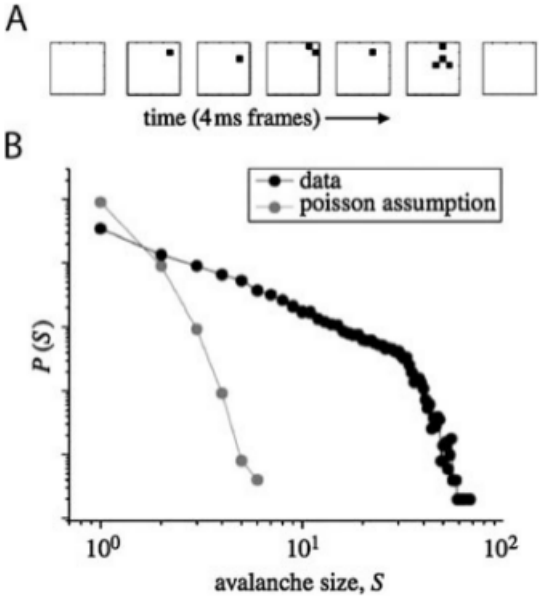


Figure 3.8: The distribution of avalanche sizes follows a power law. A. Sample avalanche propagating on the 8×8 multielectrode array. B. Probability distribution of avalanche sizes (measured in number of electrode) in log-log space. The distribution follows a power-law with a cutoff set by the size of the array. [2]

3.1.6 A Special Living System: The Brain

The brain may result as the most striking example of this universal tendency of living systems. In fact, owing to a "plastic" architecture shaped by evolution towards self-organized-criticality, brain may have gained its unmatched information processing capabilities. Neurons are microscopic processing elements forming complex networks wired through cellular processes (i.e. axons and dendrites) and specialized contacts transmitting information: the synapses. Amazingly, neuronal connections undergo "birth" and "death" as well as strengthening and weakening throughout a selection process, reconfiguring connectivity in a self-organizing manner and allowing the networked population of neuronal processors to adapt responses to the ever changing environment.

Throughout the nervous system of almost all animals, neurons communicate with one another with discrete, stereotyped electrical pulses called action potentials or spikes. If we look in a brief window of time $\Delta\tau$, the activity of a neuron is binary: a neuron either spikes ($\sigma_i = +1$) or it does not ($\sigma_i = -1, 0$).
In this notation the binary string or "spike word" $\bar{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ entirely describes the spiking activity of a network of N neurons, and the probability distribution $P(\bar{\sigma})$ over all 2^N possible spiking states describes the correlation structure of the network, as well as defining the "vocabulary" that the network has at its disposal to use in representing sensations, thoughts, memories or actions.
Neurons in the brain interact with each other in a heterogeneous and asymmetric way producing complex neuronal activity. The dynamics is given by the equations of the network, and it exhibits fixed-point behavior, limit cycles or high-dimensional chaos.
In these networks, connections between binary neurons are independently drawn from an identical distribution, and the state of a network is updated simultaneously in discrete time

steps without thermal noise. Every initial configuration must evolve into an attractor, which is a fixed point or a limit cycle. The typical length of the cycles was observed to grow exponentially with the number of neurons N (such kinds of cycles are called chaotic attractors), and the total number of attractors increases linearly with N.
These quantities were also analytically evaluated based on an empirical assumption that the dynamics of the system loses memory of its nonimmediate past.

Consider randomly connected neuronal networks of N neurons (units). Each unit interacts with all the other units with an asymmetric coupling: we use J_{ij} to represent the coupling strength from unit j to i, it is independent of J_{ji} and they follow the same Gaussian distribution $N(0, \frac{1}{N})$.
The state of neuron i $(i = 1, \ldots, n)$ at time $t + 1$ is determined by the following equation:

$$\sigma_i(t+1) = sgn\left[h_i(t)\right] = sgn\left[\sum_{j=1}^{N} J_{ij}\sigma_j(t)\right] = \begin{cases} +1 & (activate\ state) \\ \\ -1 & (silent\ state) \end{cases}$$

Sampling all the 2^N words is of course impractical, an important observation is that correlations between any two neurons typically are weak: the correlation coefficient between σ_i and $\sigma_{j\neq i}$ is on the order of 0.1 or less. It is tempting to conclude that neurons are approximately independent, and there are no interesting collective effects. Larger groups of neurons spike simultaneously much more frequently than would be expected if spiking were independent in every cell.

It might be that there are specific sub-circuits in the network that link special groups of many cells, alternatively, the network could be statistically homogeneous and simultaneous spiking of many cells emerge as a collective effect.
An important hint is that the neuronal correlations are weak and widespread, so that any two neurons that plausibly are involved in the same task are equally likely to have a significant correlation.

3.1.7 The vertebrate retina

The retina is an ideal place in which to test ideas about correlated activity, because it is possible to make long and stable recordings of many retinal ganglion cells—the output cells of the retina, whose axons bundle together to form the optic nerve—as they respond to natural visual stimuli.
Because the retina is approximately flat, one can record from the output layer of cells by placing a piece of the retina on an array of electrodes that have been patterned onto to a glass slide, using conventional methods of microfabrication. Such experiments routinely allow measurements on ~ 100 neurons.
The average rate at which neuron i generates spikes is given by

$$\bar{r}_i = \langle (1 + \sigma_1)/2 \rangle / \Delta\tau \tag{3.12}$$

so that knowing the average rates is the same as knowing the local magnetizations σ_i. A consistent model with these assumptions is a model of independently firing cells:

$$P_1(\bar{\sigma}) = \prod_i p_i(\sigma_i) = \frac{\exp\left[\sum_i h_i\sigma_i\right]}{Z} \tag{3.13}$$

where h_i is the Lagrange multiplier associated to the average observable σ_i.
For example, in a retina stimulated by natural movies, the distribution of the total number of spikes $K = \sum_{i=1}^{N}(1+\sigma_i)/2$ is observed to be approximately exponential $[P(K) \simeq e^{-K/K}]$, while an independent model predicts Gaussian tails throughout the relevant range of K.
As the first step beyond an independent model, one can look for the maximum entropy distribution that is consistent not only with σ_i , but also with pairwise correlation functions between neurons $\langle \sigma_i \sigma_j \rangle$. The distribution then takes a familiar form:

$$P_2(\bar{\sigma}) = \frac{1}{Z} e^{-E(\bar{\sigma})} \quad E(\bar{\sigma}) = -\sum_{i=1}^{N} h_i \sigma_i - \sum_{i<j} J_{ij} \sigma_i \sigma_j \tag{3.14}$$

where J_{ij} is the Lagrange multiplier associated to $\langle \sigma_i \sigma_j \rangle$.

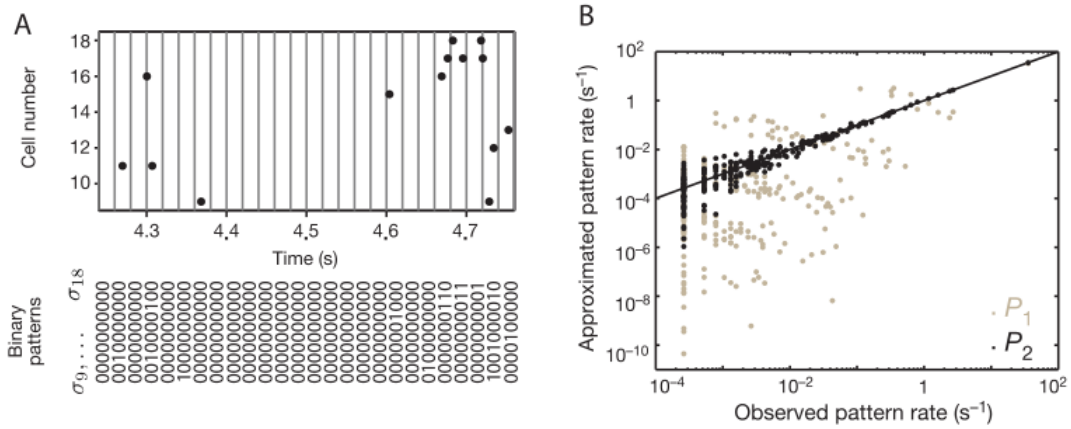


Figure 3.9: The Ising model greatly improves the prediction of retinal activity over the independent model. A. Neuronal activity is summarized by a binary word $\bar{\sigma}$ obtained by binning spikes into 20 m windows. B. The frequencies of all spike words $\bar{\sigma}$ of a subnetwork of $N = 10$ neurons are compared between the experiment (x axis) and the prediction (y axis) of the independent model (gray dots) and the maximum entropy model with pairwise interactions (black dots). The straight line represents identity.[2]

Remarkably, this model is mathematically equivalent to a disordered Ising model, where h_i are external local fields, and J_{ij} exchange couplings. Ising models were first introduced by Hopfield in the context of neuronal networks to describe associative memory.
Armed with an explicit model for the whole network, one can explore its thermodynamics. The introduction of a fictitious temperature T corresponds to a global rescaling of the fitting parameters, $h_i \to h_i/(k_B T)$, $J_{ij} \to J_{ij}/(k_B T)$.
The heat capacity versus temperature is found to be more and more sharply peaked around the operating temperature $k_B T = 1$ as one increases the network size N. Criticality could be diagnosed directly from the distribution of pairwise correlations, rather than their precise arrangement across cells. More concretely, it gives us a path to simulate what we expect to see from larger networks, assuming that the cells that have been recorded from in this experiment are typical of the larger population of cells in the neighborhood.

The result for $N = 120$ thus obtained (i.e. from fitting the Ising model to an artificial network of random correlations with the same statistics as the experiment) is an even clearer demonstration that the system is operating near a critical point in its parameter space, as shown by the peak in specific heat getting closer to the natural temperature $k_B T = 1$, shown in the top curve of Fig.3.10. This diverging heat capacity is further evidence that the system is near a critical point, but one might be worried that this is an artifact of the model or of the fitting procedure.

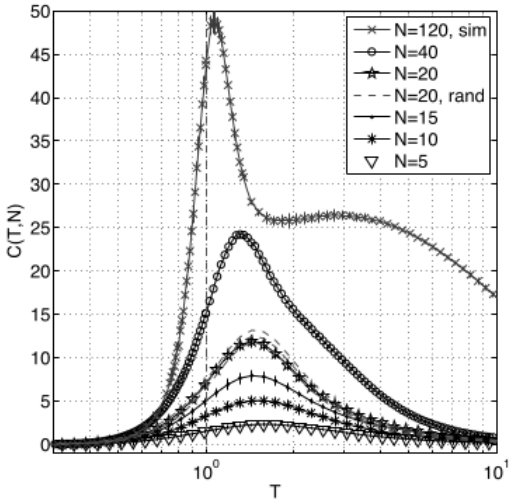


Figure 3.10: Divergence of the heat capacity is a classical signature of criticality. This plot represents the heat capacity versus temperature for Ising models of retinal activity for increasing population sizes N. The "$N = 20$, rand," and $N = 120$ curves were obtained by inferring Ising models for fictitious networks whose correlations were randomly drawn from real data. Error bars show the standard deviation when choosing different subsets of N neurons among the 40 available. [2]

3.2 Against Criticality

In this section we view some articles that support the hypothesis that a particular type of living system, the brain, is not poised at a critical state. We start with the work of Jonathan Touboul and Alain Destexhe [4] which will inspire the model that we build in the second part of the thesis. The principal argument is that, although a lot of living systems show power law distribution, they are not critical systems and power law distributions are fictitious clues of criticality.

3.2.1 Power-law statistics and universal scaling in the absence of criticality

Now we consider the work of Jonathan Touboul and Alain Destexhe [4], where they demonstrate that there can exist power-law distributions also in absence of criticality.
Criticality, whether self-organized (SOC) or the result of fine tuning is usually identified by a typical scale invariance or power-law scaling. The relationship between power-laws in nature and criticality became popular after the seminal work of Bak, Tang and Wiesenfeld on the Abelian sandpile model [44]: there, it is proposed that systems can self-organize to remain poised at their critical state and that this may be a universal explanation for the ubiquity of power-law scalings in nature [45]. In neuroscience, the theory that neuronal networks operate at a phase transition would not only be a mathematical curiosity, but could have important consequences in neural coding: a scale-invariant neuronal spiking activity could reveal the presence of long-range correlations in the system [46] and optimal encoding of information [47].

The first empirical evidence that neuronal networks may operate at criticality was provided by analysis of the activity in neuronal cultures in vitro, which display bursts of activity

separated by silences. These bursts were seen as "neuronal avalanches", and were apparently distributed as a power-law with a slope close to $-3/2$.

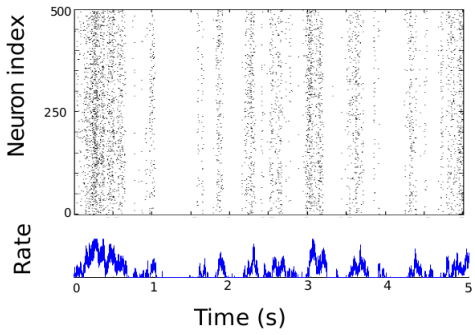


Figure 3.11: Neuronal activity in the independent Poisson model with Ornstein-Uhlenbeck firing rate ($\alpha = \sigma = 1$). [4]

These statistics were sometimes interpreted as the hallmark of criticality in analogy with canonical statistical physics models. In this work, Jonathan Touboul and Alain Destexhe show that networks with inhibition do not need to be at criticality to have power-law avalanche statistics as reported in experiments, but is rather a generic macroscopic feature of such networks related to fundamental properties of large interacting systems with inhibition. In this work they analyzed statistical distributions of neuronal avalanches. They proved

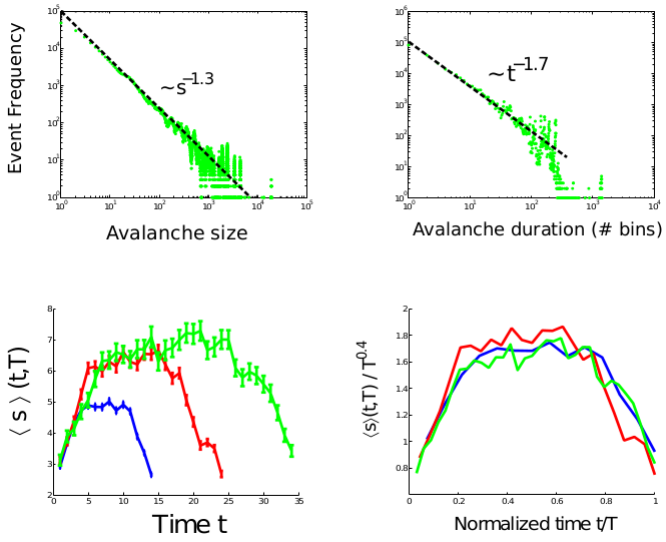


Figure 3.12: Avalanches statistics in the independent Poisson model with Ornstein-Uhlenbeck firing rate ($\alpha = \sigma = 1$). Apparent power-law scalings, together with scale invariance of avalanches shapes.[4]

that, using a (non critical) Poisson model of neuronal network with Ornstein-Uhlenbeck firing rate ($\alpha = \sigma = 1$, see Chapter 5) appear power-law scalings, together with scale invariance of avalanches shapes.

The model is better described in Chapter 5, they simulated a neuronal network of 2000 neurons and investigated the collective statistics of independent realizations of Poisson processes with that rate. While the firing is an inhomogeneous Poisson process, macroscopic statistics show power-law distributions for the size (with exponent $\tau = 1.3$) and for the duration ($\alpha = 1.7$).

At this point, they conclude that power-law statistics of avalanche duration and size, as well as collapse of avalanche shape, may arise in non-critical systems, they even appear in networks in which firing times are statistically independent. Therefore, they do not necessarily reveal how neurons interact, but rather the presence of a hidden dependence of all neurons on the same fluctuating function.
They conclude that networks of neurons within synchronous irregular (SI) regimes may naturally display power-law statistics and universal scaling functions due to the combination of (i) an emergent irregular collective activity in a large-scale system, and (ii) the power-law statistics of such large-scale systems, due to molecular chaos.

3.2.2 Neural Net Model

One model analyzed by David A. Kessler and Herbert Levine [15] in August 2015 is inspired by the dynamic competition between different species, used as the main nonlinear ingredient of their ecology model. They were thus motivated to turn out a model with a similar interaction in a different context, a set of neurons which are mutually inhibitory and are driven by external sources. The inhibition matrix $J_{i,j}$ is constructed from random elements of average size unity (distributed uniformly between 0 and 2), with density ρ, with all other elements set to zero. The voltage level on a given neuron decays exponentially with time constant τ. The external input, V_i^I is uniformly distributed at each time slice dt

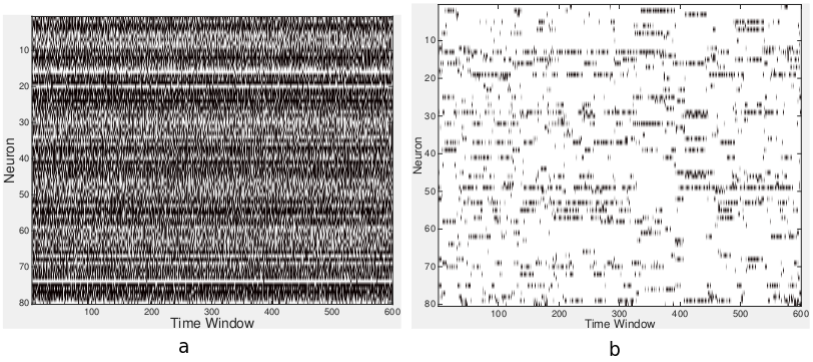


Figure 3.13: Sample firing patterns from our neuronal net model with $N = 80$ neurons. Left: Pattern with constant threshold $V_c = 19.67$. The time window was $T_W = T_p = 83$. Right: Pattern with dynamical thresholds, governed by the parameters $f_T = 0.85$, $\gamma_T = 0.02$. Due to the lower threshold of the adapted neurons, the time window was taken as $T_W = 40$. For both models, $C = 1$, $\sigma = 0.7$, $\sigma I = 0.02$, $\tau = 20$.[15]

between $dt(1 \pm 3\sigma_I)$. The threshold for a neuron firing is given initially by

$$V_c = \tau \left(1 - e^{-T_p/\tau}\right) \tag{3.15}$$

so that in the absence of noise, every neuron would fire with period T_p. Thus, the update rule for V_i is

$$V_i(t + dt) = e^{-dt/\tau} V + V_i^I \tag{3.16}$$

At this point, suprathreshold neurons fire, and are reset to 0. The firing neurons then reduce the voltage on the neurons they are connected to according to

$$V_i = V_i - C \sum_{j \in fire} J_{i,j} \tag{3.17}$$

A pictorial view of the dynamics is given in Fig.3.13a. Essentially, the strong inhibition and the lack of any long-term memory that a neuron has been active serves to create an extremely chaotic process; in the ecology model the role of memory is played by the size of a species population which if large prevents the species from rapidly disappearing.

To mimic this effect, we added a feedback between firing and threshold. When a neuron fires, its threshold for future firing is lowered, by a factor f_T. The threshold then relaxes exponentially back to its nominal value, V_c , with a relaxation rate of γ_T. One could imagine accomplishing analogous changes by altering synapses, i.e. weakening the synaptic inhibition for any neuron that fires. Once this change is implemented, the dynamics stabilizes to the form seen in Fig.3.13b.
The period of firing of a neuron is exponentially sensitive to its threshold when the threshold is close to its critical value of τ. David A. Kessler and Herbert Levine then tabulate the statistics of firing patterns. Their first set of runs is with $N = 80$ neurons and parameters $\rho = 1$, $\tau = 20$, $T_p = 83$, $C = 1$, $\sigma_I = 0.02$. In Fig.3.14, is reported the complementary cumulative distribution function. We can observe an almost perfect power law form, scaling with

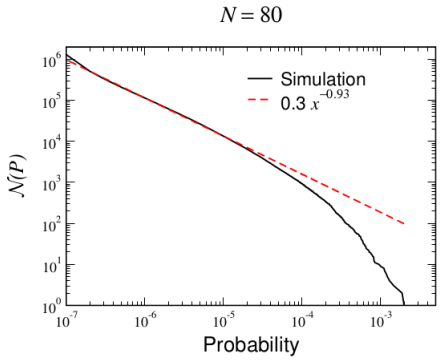


Figure 3.14: Complementary cumulative distribution function for $N = 80$. Parameters are the same as in the simulation shown on the right side of Fig.3.13.[15]

exponent close to -1. This would guarantee that any attempt to model the thermodynamics of the system with an equilibrium Hamiltonian would "discover" that the system looks like it had been tuned to be very close to a critical point. Again however, this state exists over a broad range of parameters and does not require any fine-tuning. A similar tabulation for the non-adaptive model shows that the system thermodynamics is completely dominated by a huge number of very low occupancy states.

3.2.3 Linear regression and power laws

The most common approach for testing empirical data against a hypothesized power-law distribution is to observe that the power law $p(x) \sim x^{-\alpha}$ implies the linear form

$$\log p(x) = \alpha \log x + c \tag{3.18}$$

The probability density $p(x)$ can be estimated by constructing a histogram of the data (or alternatively one can construct the cumulative distribution function by a simple rank ordering

of the data) and the resulting function can then be fitted to the linear form by least-squares linear regression. The slope of the fit is interpreted as the estimate $\hat{\alpha}$ of the scaling parameter.
Although this procedure appears frequently in the literature there are several problems with it. The estimates of the slope are subject to systematic and potentially large errors, but there are a number of other serious problems as well. First, errors are hard to estimate because they are not well-described by the usual regression formulas, which are based on assumptions that do not apply in this case. For continuous distributions used for discrete data, this problem can be exacerbated by the choice of binning scheme used to construct the histogram, which introduces an additional set of free parameters. Second, a fit to a power-law distribution can account for a large fraction of the variance even when the fitted data do not follow a power law. Third, the fits extracted by regression methods usually do not satisfy basic requirements on probability distributions, such as normalization, and hence cannot be correct.

Many standard packages exist that can perform this kind of fitting, provide estimates and standard errors for the slope, and calculate the fraction r^2 of variance accounted for by the fitted line, which is taken as an indicator of the quality of the fit.
A data set has n values marked $y_1 \dots y_n$ (collectively known as y_i), each associated with a predicted (or modeled) value $f_1 \dots f_n$ (known as f_i, or sometimes $\hat{y}_i$). If $\bar{y}$ is the mean of the observed data:

$$r^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3.19}$$

If our data are truly drawn from a power-law distribution and n is large, then the probability of getting a low r^2 in a straight-line fit is small, so a low value of r^2 can be used to reject the power-law hypothesis. Distributions that are nothing like a power law can appear to follow a power law for small samples and some, like the log-normal, can approximate a power law closely over many orders of magnitude, resulting in high values of r^2. Though a low r^2 is informative, in practice we rarely see a low r^2, regardless of the actual form of the distribution, so that the value of r^2 tells us little. In the terminology of statistical theory, the value of r^2 has very little power as a hypothesis test because the probability of successfully detecting a violation of the power-law assumption is low.

Maximum likelihood estimators for the power law

In the case of continuous data the maximum likelihood estimator for the scaling parameter, first derived by Muniruzzaman in 1957 [18], is equivalent to the well-known Hill estimator [19]. Consider the continuous power-law distribution,

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha} \tag{3.20}$$

where α is the scaling parameter and x_{min} is the minimum value at which power-law behavior holds. Given a data set containing n observations $x_i \geq x_{min}$, we would like to know the value of α for the power-law model that is most likely to have generated our data. The probability that the data were drawn from the model is proportional to

$$p(x|\alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} \tag{3.21}$$

This probability is called the *likelihood* of the data given the model. The data are most likely to have been generated by the model with scaling parameter α that maximizes this

function. Commonly we actually work with the logarithm $\mathcal{L}$ of the likelihood, which has its maximum in the same place:

$$\mathcal{L} = \log p(x|\alpha) = \log \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} = \tag{3.22}$$

$$= n \log(\alpha - 1) - n \log x_{min} - \alpha \sum_{i=1}^{n} \log \frac{x_i}{x_{min}} \tag{3.23}$$

Setting $\partial \mathcal{L} / \partial \alpha = 0$ and solving for α, we obtain the maximum likelihood estimate or MLE for the scaling parameter:

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^{n} \log \frac{x_i}{x_{min}} \right]^{-1} \tag{3.24}$$

Formal results

There are a number of formal results in mechanical statistics that motivate and support the use of the MLE:

Theorem 3.2.1. *Under mild regularity conditions, if the data are independent, identically-distributed draws from a distribution with parameter α, then as the sample size $n \to \infty$, $\hat{\alpha} \to \alpha$ almost surely.*

Proposition 3.2.2. *The maximum likelihood estimator $\hat{\alpha}$ of the continuous power law converges almost surely on the true α.*

Proof. It is easily verified that $\log(x/x_{min})$ has an exponential distribution with rate $\alpha - 1$. By the strong law of large numbers, therefore, $\frac{1}{n} \sum_{i=1}^{n} \log \frac{x_i}{x_{min}}$ converges almost surely on the expectation value of $\log(x/xmin)$, which is $(\alpha - 1)^{-1}$. □

Chapter 4

A "Poisson Network"

In this chapter, following by the work of Johnatan Touboul and Alain Destexhe [4], we analyze if a random non-interacting neuronal network model can generate emergent properties that are typical of critical systems (e.g., power law distributions). In this model, every neuron spikes accordingly to a nonhomogeneous Poisson process with a firing rate that is generated from an exponential distribution. The statistic of activity avalanches then computed.

We first introduce some definitions and theorems that are used in the simulations of the Poisson network model. Later, we describe the neural network model in detail.

4.1 The Exponential Distribution

Definition 2. A continuous random variable X is said to have an exponential distribution with parameter $\lambda,\ \lambda > 0$, if its probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq o \\ \\ 0 & x < 0 \end{cases} \tag{4.1}$$

or, equivalently, if its CDF is given by

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \begin{cases} 1 - e^{-\lambda x} & x \geq o \\ \\ 0 & x < 0 \end{cases} \tag{4.2}$$

The mean of the exponential distribution, $E[X]$, is given by

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x)\, dx = \frac{1}{\lambda} \tag{4.3}$$

In general, the moment generating function $\phi(t)$ of the exponential distribution is given by

$$\phi(t) = \mathbb{E}[e^{tx}] = \int_{0}^{+\infty} e^{tx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - t} \tag{4.4}$$

All the moments of X can now be obtained by differentiating Eq.(4.4). For example,

$$\mathbb{E}[X^2] = \left(\frac{d^2}{dt^2} \phi(t) \right)_{t=0} = \frac{2}{\lambda^2} \tag{4.5}$$

Consequently, $Var(X) = 1/\lambda^2$.

Definition 3. A random variable X is said to be without memory, or memoryless, if

$$P\{X > s+t | X > t\} = P\{X > s\} \quad \forall\, s,t \geq 0 \tag{4.6}$$

If we think of X as being the lifetime of some instrument, then Eq.(4.6) states that the probability that the instrument lives for at least $s+t$ hours given that it has survived t hours is the same as the initial probability that it lives for at least s hours. In other words, if the instrument is alive at time t, then the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution; that is, the instrument does not remember that it has already been in use for a time t.
The condition in Eq.(4.6) is equivalent to

$$\frac{P\{X > s+t, X > t\}}{P\{X > t\}} = P\{X > s\} \tag{4.7}$$

or

$$P\{X > s+t, X > t\} = P\{X > t\}P\{X > s\} \tag{4.8}$$

This result is satisfied when X is exponentially distributed (for $e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}$), it follows that exponentially distributed random variables are memoryless.

Let $X_1, \ldots, X_n$ be independent and identically distributed exponential random variables having mean $1/\lambda$. It follows that $X_1, \ldots, X_n$ has a gamma distribution with parameters n and λ. Let us now give a second verification of this result by using mathematical induction. Because there is nothing to prove when $n = 1$, let us start by assuming that $X_1, \ldots, X_{n-1}$ has density given by

$$f_{X_1+\cdots+X_{n-1}}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-2}}{(n-2)!} \tag{4.9}$$

Hence,

$$f_{X_1+\cdots+X_{n-1}+X_n}(t) = \int_0^\infty f_{X_n}(t-s) f_{X_1,\ldots,X_{n-1}}(s)\, ds = \tag{4.10}$$

$$= \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s}\frac{(\lambda s)^{n-2}}{(n-2)!} = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} \tag{4.11}$$

which proves the result.
Another useful calculation is to determine the probability that one exponential random variable is smaller than another. That is, suppose that X_1 and X_2 are independent exponential random variables with respective means $1/\lambda_1$ and $1/\lambda_2$; what is $P\{X_1 < X_2\}$? This probability is easily calculated by conditioning on X_1:

$$P\{X_1 < X_2\} = \int_0^\infty P\{X_1 < X_2 | X_1 = x\}\lambda_1 e^{-\lambda_1 x}\, dx = \tag{4.12}$$

$$= \int_0^\infty P\{X_1 < X_2\}\lambda_1 e^{-\lambda_1 x}\, dx = \int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x}\, dx = \frac{\lambda_1}{\lambda_1+\lambda_2} \tag{4.13}$$

4.2 The Poisson Process

We define now the Poisson process, in its homogeneous and nonhomogeneous forms. This type of process will be used to describe the firing of neurons of the neuronal network.

4.2.1 Counting Processes

A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of "events" that occur by time t. From its definition we see that for a counting process $N(t)$ must satisfy:

1. $N(t) \geq 0$.
2. $N(t)$ is integer valued.
3. If $s < t$, then $N(s) < N(t)$.
4. For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

A counting process is said to possess independent increments if the numbers of events that occur in disjoint time intervals are independent. A counting process is said to possess stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. In other words, the process has stationary increments if the number of events in the interval $(s, s+t)$ has the same distribution for all s.

4.2.2 Definition of the Poisson Process

One of the most important counting processes is the Poisson process, which is defined as follows:

Definition 4. The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* having rate λ, $\lambda > 0$, if

- $N(0) = 0$.
- The process has independent increments.
- The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad n = 0, 1 \ldots \tag{4.14}$$

Note that it follows that a Poisson process has stationary increments and also that

$$\mathbb{E}[N(t)] = \lambda t \tag{4.15}$$

which explains why λ is called the rate of the process.

Definition 5. The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$, if

- $N(0) = 0$.
- The process has stationary and independent increments.
- $P\{N(h) = 1\} = \lambda h + o(h)$.
- $PN(h) \geq 2 = o(h)$.

It can be shown that Definition 4 and Definition 5 are equivalent.

4.3 Generalization of Poisson Process: Nonhomogeneous Poisson Process

In this section we consider a generalization of the Poisson process: the nonhomogeneous, also called the nonstationary Poisson process, which is obtained by allowing the arrival rate at time t to be a function of t.

Definition 6. The counting process $\{N(t), t \geq 0\}$ is said to be a *nonhomogeneous Poisson process with intensity function* $\lambda(t)$, $t \geq 0$, if

- $N(0) = 0$.
- $\{N(t), t \geq 0\}$ has independent increments.
- $P\{N(t+h) - N(t) \geq 2\} = o(h)$.
- $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$.

Time sampling an ordinary Poisson process generates a nonhomogeneous Poisson process. That is, let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ, and suppose that an event occurring at time t is, independently of what has occurred prior to t, counted with probability $p(t)$. With $N_c(t)$ denoting the number of counted events by time t, the counting process $\{N_c(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t) = \lambda p(t)$. This is verified by noting that $\{N_c(t), t \geq 0\}$ satisfies the nonhomogeneous Poisson process axioms.
Every nonhomogeneous Poisson process with a bounded intensity function can be thought of as being a time sampling of a Poisson process. To show this, we start by showing that the superposition of two independent nonhomogeneous Poisson processes remains a nonhomogeneous Poisson process.

Proposition 4.3.1. *Let* $\{N(t), t \geq 0\}$, *and* $\{M(t), t \geq 0\}$, *be independent nonhomogeneous Poisson processes, with respective intensity functions* $\lambda(t)$ *and* $\mu(t)$, *and let* $N^*(t) = N(t) + M(t)$. *Then, the following are true.*

1. $\{N^*(t), t \geq 0\}$ *is a nonhomogeneous Poisson process with intensity function* $\lambda(t)+\mu(t)$.
2. *Given that an event of the* $\{N^*(t)\}$ *process occurs at time* t *then, independent of what occurred prior to* t, *the event at* t *was from the* $\{N(t)\}$ *process with probability* $\frac{\lambda(t)}{\lambda(t)+\mu(t)}$.

4.4 How to Generate Poisson Processes

A simple and relatively efficient method for simulating one-dimensional and two-dimensional nonhomogeneous Poisson processes is presented.
The method is applicable for any rate function and is based on controlled deletion of points in a Poisson process whose rate function dominates the given rate function. As seen in the previous section, the one-dimensional nonhomogeneous Poisson process has the characteristic properties that the numbers of points in any finite set of nonoverlapping intervals are mutually independent random variables, and that the number of points in any interval has a Poisson distribution.
Simulation of a nonhomogeneous Poisson process with general rate function $\lambda(t)$ of a nonhomogeneous Poisson process with rate function $\lambda^*(t) \geq \lambda(t)$. The basic result is the following theorem.

Theorem 4.4.1. *Consider a one-dimensional nonhomogeneous Poisson process* $\{N^*(t) : t \geq 0\}$ *with rate function* $\lambda^*(t)$*, so that the number of points,* $N^*(t_0)$*, in a fixed interval* $(o, t_0]$ *has a poisson distribution with parameter* $\mu_0^* = \lambda^*(t_0) - \lambda^*(0)$*. Let* $\{t_i^*,\ i = 1 \dots N^*(t_0)\}$ *be the points of the process in the interval* $(o, t_0]$*. Suppose that for* $t \in [0, t_0],\ \lambda(t) \leq \lambda^*(t)$*. For* $i = 1, 2, \dots, n$ *delete the point* t_i^* *with probability* $1 - \lambda^*(t_i)/\lambda^*(t_i^*)$*; then the remaining points form a nonhomogeneous Poisson process* $\{N(t) : t \geq 0\}$ *with rate function* $\lambda(t)$ *in the interval* $(o, t_0]$*.*

Algorithm

1. Generate points in the nonhomogeneous Poisson process $\{N^*(t)\}$ with rate function $\lambda^*(t)$ in the fixed interval $(o, t_0]$. If the number of points generated, n^*, is such that $n^* = 0$, exit; there are no points in the process $\{N(t)\}$.
2. Denote the (ordered) points by $T_1^*, T_2^*, \dots, T_n^*$. Set $i = 1$ and $k = 0$.
3. Generate U_i, uniformly distributed between 0 and 1. If $U_i \leq \lambda(T_i^*)/\lambda^*(T_i^*)$, set k equal to $k + 1$ and $X_k = T_i^*$.
4. Set i equal to $i + 1$. If $i \leq n^*$, go to 3.
5. Return $T_1, T_2, \dots, T_n$ where $n = k$, and also n.

4.5 The Poisson Network Model

The first step of this master's thesis is to analyze a neuronal network in the hypothesis of independent neurons: every neuron spikes with exponential distribution and the rate $\lambda(t)$ is generated from an exponential distribution of rate γ.
To proceed we have to generate $\lambda(t)$ directly from the following exponential distribution:

$$p(\lambda) = \gamma e^{-\gamma\lambda} \tag{4.16}$$

To draw λ_is from this type of distribution we have to calculate the cumulative distribution $P(\lambda)$:

$$P(\lambda) = \int_0^{\lambda} \gamma e^{-\gamma x}\, dx = 1 - e^{-\gamma\lambda} \tag{4.17}$$

Using this result and drawing a set of random numbers $\{u,\ P_>(\lambda) = u\}$ directly from a uniform distribution, the set of λs can be calculated as follows:

$$\lambda = -\frac{1}{\gamma}\ln(1 - u) \quad u \in \mathcal{U}[0, 1] \tag{4.18}$$

that is equivalent to

$$\lambda = -\frac{1}{\gamma}\ln(u) \quad u \in \mathcal{U}[0,1] \tag{4.19}$$

because $(1-u)$ is distributed exactly as u.
Every λ can be calculated after discretized the interval $[0;T]$ with dt such that $N^*dt = T$. Using this method, the typical pattern obtained for $\lambda(t)$, $t \in [0;1]$, is reported in Fig.4.1.

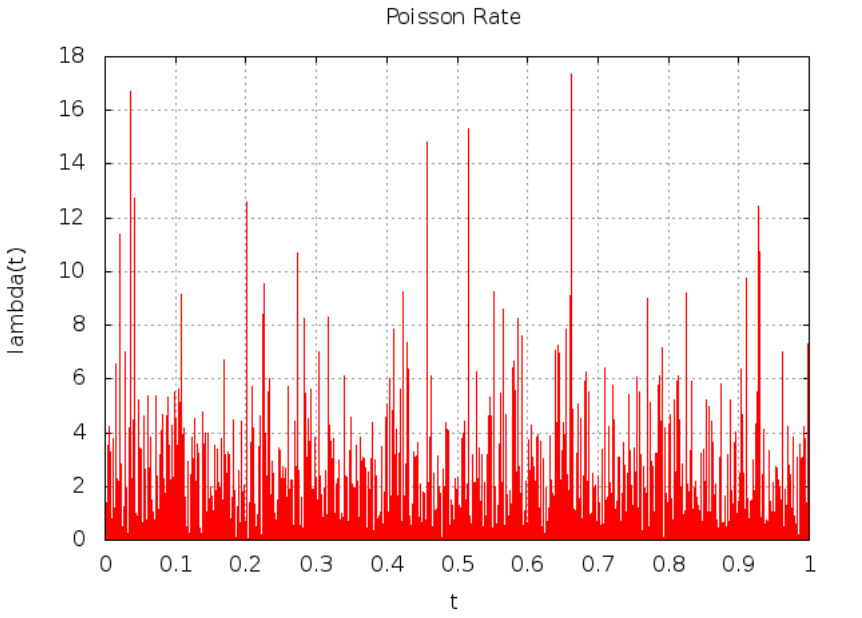


Figure 4.1: Tipical pattern $\lambda(t)$ obtained from an exponential distribution with $\gamma = 0.5$. The binning in the interval $[0;1]$ is done with $dt = 0.001$.

Using this rate $\lambda(t)$ as the rate of spiking neurons as described in the next section, we analyze the principal characteristics of the neuronal activity.

4.6 The Nonhomogeneous Spiking Process

After having generated the rate $\lambda(t)$ that is universal for every neuron which we are going to analyze, it is fundamental to define what type of process describes the activity of each neuron.
This process is the nonhomogeneous Poisson process, that is already described in this chapter. For our case, implementing of this process is simple. Following the general method of Poisson process, every neuron spikes following an exponential distribution with rate λ_{max}, where $\lambda_{max} \equiv Max\{\lambda(t), t \in [0;T]\}$, so for every neuron of the network this Poisson process is computed following an exponential distribution of rate λ_{max}:

$$p(t) = \lambda_{max} e^{-\lambda_{max} t} \tag{4.20}$$

that is, the probability that a spike occurs after a time interval t given the mean rate λ_{max}. Using Eq.(4.20), the process for every neuron is described as follows:

$$t_{i+1} = t_i - \frac{1}{\lambda_{max}}\ln(u) \quad u \in \mathcal{U}[0,1] \quad i = 0, 1 \ldots \tag{4.21}$$

with $t_{i=0} = 0$.
For every time step is applied the thinning method to generate the nonhomogeneous Poisson process: for a generic t_j, it is evaluated the quantity $\lambda(t_j)/\lambda_{max}$. After drawing $u \in \mathcal{U}[0;1]$, if $\lambda(t_j)/\lambda_{max} \geq u$ the spike event is accepted, otherwise is rejected.

How to evaluate the quantity $\lambda(t_j)$ for $t_j \in [t_i; t_{i+1}]$, $i \in \mathbb{N}$?
The method is simple, after having discretized the interval $[0;T]$, every $\lambda(t_i)$, $i = 0, 1, \ldots, N$, corresponds to the i-th bin such that $t_i = i^* dt$. Then, after identified $\lambda_i \equiv \lambda(t_i)$, the correspondent $\lambda(t_j)$ for $j \in [t_i; t_{i+1}]$ is calculated using the straight line passing through two points formula:

$$\frac{\lambda - \lambda_i}{\lambda_{i+1} - \lambda_i} = \frac{t - t_i}{t_{i+1} - t_i} \tag{4.22}$$

So, equivalently,

$$\lambda = \lambda_i + \frac{\lambda_{i+1} - \lambda_i}{t_{i+1} - t_i}(t - t_i) \tag{4.23}$$

To find a generic $\lambda(t_j)$, remembering the system's discretization:

$$\lambda(t_j) = \lambda_i + \frac{\lambda_{i+1} - \lambda_i}{dt}(t_j - t_i) \tag{4.24}$$

Using this result, it can be estimated a generic $\lambda(t_j)$ for every $t_j \in [0;T]$.

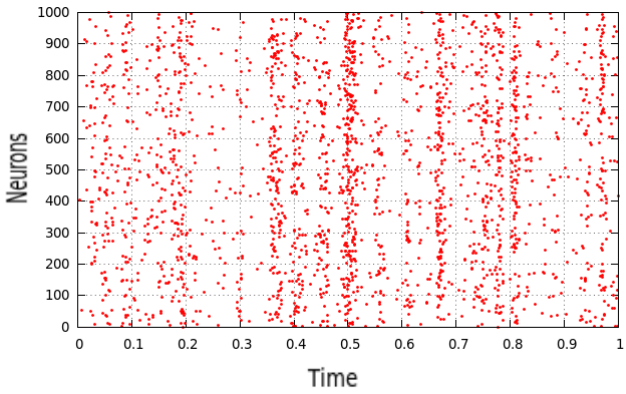


Figure 4.2: Tipical pattern for 1000 neurons spiking using a nonhomogeneous Poisson process with $\gamma = 0.5,\ dt = 0.001$.

4.7 Avalanche Statistics in a Discrete Poisson Process

In the particular case of *discrete Poisson process* which we are going to analyze, the thinning method can be replaced by a simpler method that uses the definition of the Poisson process. Anyway, the thinning method will be used in the following chapters in the case of a rate generated by a Ornstein-Uhlenbeck process.
In this work we have considered both the continuous and the discrete Poisson process. In the first simulation we used the thinning algorithm previously described, and in the discrete Poisson process for every bin i (or time step $t_i = i^* dt$) the neuron is allowed to spike if and only if $\lambda_i dt \geq u$, $u \in \mathcal{U}[0;1]$. The results of these two models were identical in both cases, for this reason in this chapter we analyse the simpler process, that is the discrete Poisson

process.

Given the pattern in Fig.4.3, one can project the neuronal activity on the time axis, and

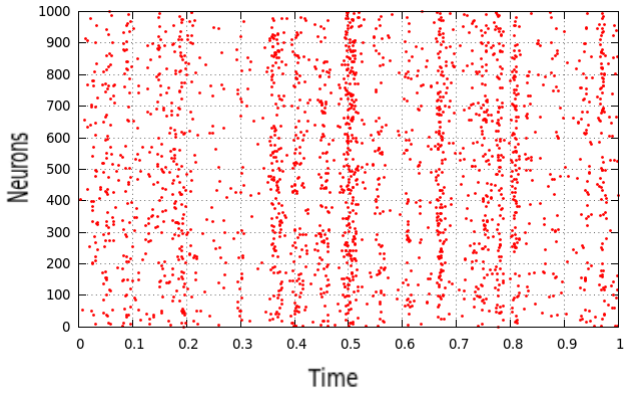


Figure 4.3: Tipical pattern for 500 neurons spiking using a nonhomogeneous discrete Poisson process with $\gamma = 0.5,\ dt = 0.001$.

obtaining the so-called *neuronal activity* and the *neuronal avalanches*.

One neuronal avalanche is defined when, starting from null neuronal activity, the neuronal network admits an activity that grows and decades until the total activity returns to be null. These avalanches can be analyzed: it can be analyzed the probability to have an avalanche of time duration t and the probability to have an avalanche of size s. The time duration $t = t_{end} - t_{init}$ is the time interval between the beginning of the avalanche t_{init} and the ending t_{end} of it, whereas the size s is defined as the total number of neurons activated during the avalanche.

What type of statistics do we expect from the neuronal activity? It is surely a non critical system for as we generate it. So, in principle we do not expect power law statistics, as it is provided by analytical calculations which will be used in this chapter.

The fact that every neuronal spike is generated by an homogeneous Poisson and the rate is generated from an exponential distribution makes that do not exist correlations between the neurons. This model of uncorrelated neurons is useful to comprehend how neuronal networks behave in general without correlations, we will see that correlations are the key ingredient to generate a power law statistics in living systems.

4.7.1 Avalanche Duration

What is the probability $P_>(t)$ of having a neuronal avalanche of duration greater than t? Analyzing the pattern in Fig.4.3, one founds that the cumulative probability depends first on the Δt of the analysis of neuronal activity. What is Δt? It is the "window" with which avalanches are analyzed. This window can change significantly the distribution that we are going to analyze, obviously because bigger is the window, less it can identify smaller avalanches. The analysis of the avalanches is simple: once obtained the pattern $N(t)$ of the neuronal activity that describes how many neurons are activated in a specific time interval dt, identifying the regions where the total number of active neurons is zero it can be possible to determine the size and the duration of the avalanches.

Once identified the neuronal avalanches, we calculate the cumulative probability that one avalanche have time duration greater than t. To do this, once fixed t, we count how many avalanches have time duration smaller or equal to t (n_t) and divide this number with the

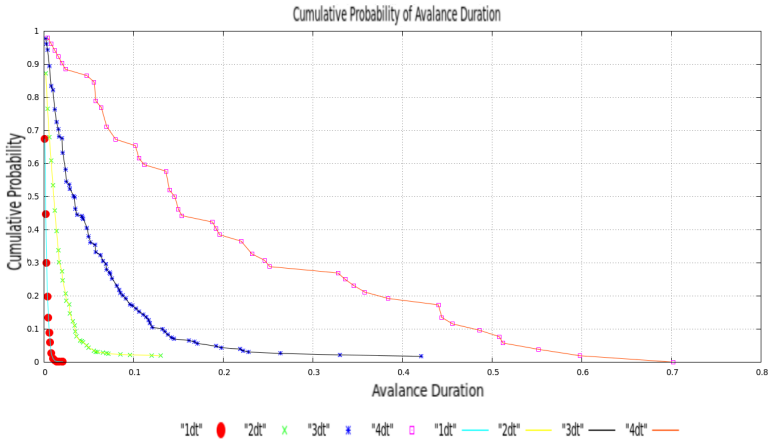


Figure 4.4: The cumulative probability of having an avalanche duration greater than $t = ndt$ for different Δt of analysis, using a nonhomogeneous discrete Poisson process with $\gamma = 0.5$, $dt = 0.001$ and 1000 neurons.

total number of avalanches N. Now we have obtained the cumulative probability that an avalanche has time duration smaller than $P_<(t)$, so to have $P_>(t)$, we take the complementary cumulative probability: $P_> = 1 - n_t/N$.

As result, the cumulative probability $P_>(t)$ of having an avalanche duration greater than t for this type of systems is reported in Fig.4.4.

The same plot in Log-y-scale is reported in Fig.4.5.

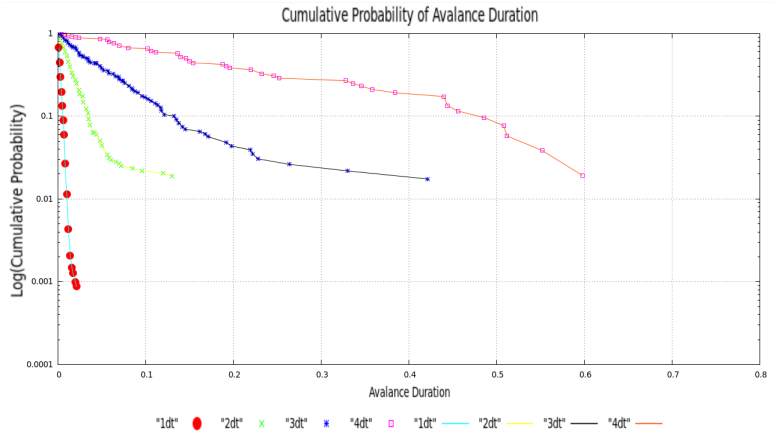


Figure 4.5: The cumulative probability of having an avalanche duration greater than $t = ndt$ in Log-y scale for $\gamma = 0.5$, $dt = 0.001$ and 1000 neurons.

We have seen that the cumulative distribution of the avalanche durations follows an exponential relation of the form $P_>(t) \simeq a^{-mt}$. In Table 4.1 are reported the various coefficients at the exponent of the distribution that depend on Δt, fitted directly from the distributions in Fig.4.4.

Note that the behavior of $P_>(t)$ depends on the Δt with which is done the analysis. We can see that m decreases when Δt increases. Analyzing the system with $\Delta t = 1dt$ or with

Δt	$1dt$	$2dt$	$3dt$	$4dt$	$5dt$	$10dt$
m	418.768 ± 14	65.7 ± 1	18.3 ± 0.3	4.66 ± 0.1	1.3 ± 0.1	0.09 ± 0.01

Table 4.1: Exponents of the exponential distribution fit $P_>(t) \simeq e^{-mt}$ of avalanche durations in the Poisson network reported in Fig.4.5.

$\Delta t = 10dt$ in general results in different behaviors of $P_>(t)$. In this case we can see that the cumulative probability is an exponential distribution and the it changes with the change of Δt. Why does it happen? Let's analyze this problem through an analytical computation.

4.7.2 Comparisons with Analytical Result

Let us consider a system composed by N neurons, each of them modeled by an independent Cox Process[1] with with intensity $\lambda(t)$. We want $\lambda(t)$ to be constant in each interval $[jdt, (j+1)dt]$, $j = 1, ...$ Thus, spikes of each neuron are a non homogeneous Poisson point process with time dependent intensity $\lambda(t) := \sum_{j=1} \lambda_j \xi_{[jdt,(j+1)dt]}$ where ξ_A is the indicator function of the set A. We work under the hypothesis that $\lambda_1, \lambda_2, ..., \lambda_n \sim \lambda$ are IID random variable with common distribution Q.
We now look to $P > (t|\lambda_1, \lambda_2 \ldots, \lambda_n)$ that is the probability of having an avalanche size of duration longer than $t = n \times dt$ given the story of λ up to time t. In what following we set $dt = \alpha/N$ and thus $n = Nt/\alpha$.
The probability that at least one neuron spikes in small interval dt is $1 - (1 - \lambda_n dt)^N$ and thus we can write

$$P_>(t|\lambda_1, \lambda_2 \ldots, \lambda_n) = [1 - (1 - \lambda_n dt)^N] P_>(t-1|\lambda_1, \lambda_2 \ldots, \lambda_n) \tag{4.25}$$

that is a recursive equation for $P_>(t)$ ($P_>(0) = 1$):

$$P_>(t|\lambda_1, \lambda_2 \ldots, \lambda_n) = [1 - (1 - \lambda_n dt)^N][1 - (1 - \lambda_{n-1} dt)^N] \ldots [1 - (1 - \lambda_1 dt)^N] \tag{4.26}$$

Therefore, marginalizing over the λs we have

$$P_>(t) = \int \prod_{i=1}^{n} dQ(\lambda_i)[1 - (1 - \lambda_i dt)^N], \tag{4.27}$$

recalling that $dt = \alpha/N$ and $n = Nt/\alpha$ we have

$$P_>(t) = \left[\int dQ(\lambda)[1 - (1 - \lambda \times \alpha/N)^N] \right]^{Nt/\alpha} \tag{4.28}$$

The limit for $N \to \infty$ is an exponential in t.
In the case of the independent Poisson process $Q(\lambda) = \gamma e^{-\gamma\lambda}$, for $\Delta t = 1dt$ the result is the following:

$$P_>(t) = \left[\int dQ(\lambda)[1 - e^{-\lambda\alpha}] \right]^{Nt/\alpha} = \left[\int_0^\infty \gamma e^{-\gamma\lambda} d\lambda(\lambda)[1 - e^{-\alpha\lambda}] \right]^{Nt/\alpha} = \tag{4.29}$$

$$= \left[1 - \frac{\gamma}{\alpha + \gamma} \right]^{Nt/\alpha} = \left[\frac{\alpha}{\alpha + \gamma} \right]^{Nt/\alpha} \tag{4.30}$$

[1] Also known as a doubly stochastic Poisson process or mixed Poisson process, is a stochastic process which is a generalization of a Poisson process where the time-dependent intensity $\lambda(t)$ is itself a stochastic process

So for the Poisson process the probability is

$$P_{>}(t) = \left[\frac{\alpha}{\alpha+\gamma}\right]^{Nt/\alpha} \tag{4.31}$$

That is an exponential in t with base $\frac{\alpha}{\alpha+\gamma}$.

The analytical result for $\gamma = 0.5$ is the following:

$$P_{>}(t) = \left[\int dQ(\lambda)[1-(1-\lambda\times\alpha/N)^N]\right]^{Nt/\alpha} \tag{4.32}$$

where $1000x = 1000t = n$ is the number of bin (time steps) of an avalanche.

Fig.4.6 shows the comparison of the analytical result $f(x)$ with the computational one in which was fixed $\Delta t = 1dt$.

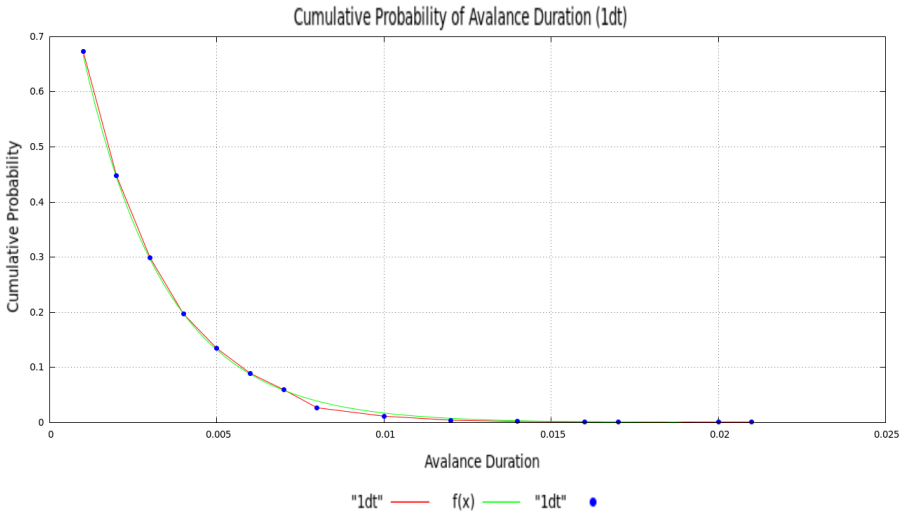


Figure 4.6: Comparison of the analytical result $f(x)$ (green line) and the computational results (blue points connected by the red line). $\gamma = 0.5$, $dt = 0.001$ and 1000 neurons.

The slope of the straight line analytically calculated $m = 1000 \times \ln(2/3) = -405.47$, compatible with that of the fit of the simulation $m_c = -418.768 \pm 13.68$.
Note that the analytical result is very similar to the computational result. To calculate the cumulative probabilities for $\Delta t = ndt$ with $n > 1$ we would have to consider the correlations between the time intervals. By the way, this is not very easy to calculate, but we can deduce from the general analytical result that the form is yet an exponential:

$$P_{>}(t) = \int \prod_{i=1}^{n} dQ(\lambda_i)[1-(1-\lambda_i dt)^N], \tag{4.33}$$

It can be seen that, although we use $\Delta t = ndt$, $n > 1$, for the analysis of the avalanche duration, we will obtain a different basis and exponent. Anyhow, in the exponent will compare the number of the bins n of the avalanche.

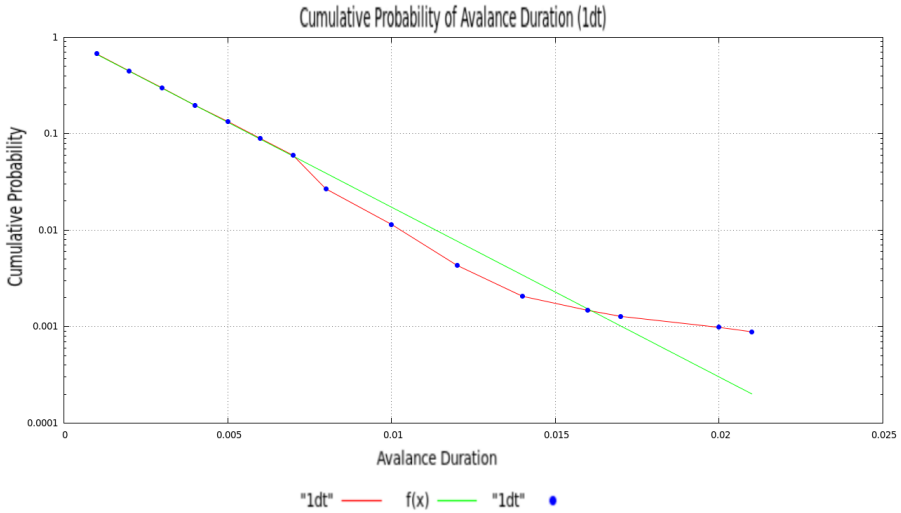


Figure 4.7: The same comparison of Fig.4.6 in Log-y-scale.

4.7.3 Mean Avalanche Duration

We want to calculate what is the mean duration $\langle t \rangle$ of an avalanche of fixed size s. To find this relation, we calculated the mean duration $\langle t \rangle (s)$ of an avalanche, for a given s:

$$\langle t \rangle (s) = \frac{1}{n} \sum_{i=1}^{n} t_i(s) \tag{4.34}$$

In Fig.4.8 we see the relationship between $\langle t \rangle (s)$ and s. The data seem to have the asymptotic power law (almost linear) behavior $\langle t \rangle (s) \sim s^a$ for big sizes with exponent $a = 0.97 \pm 0.04$.

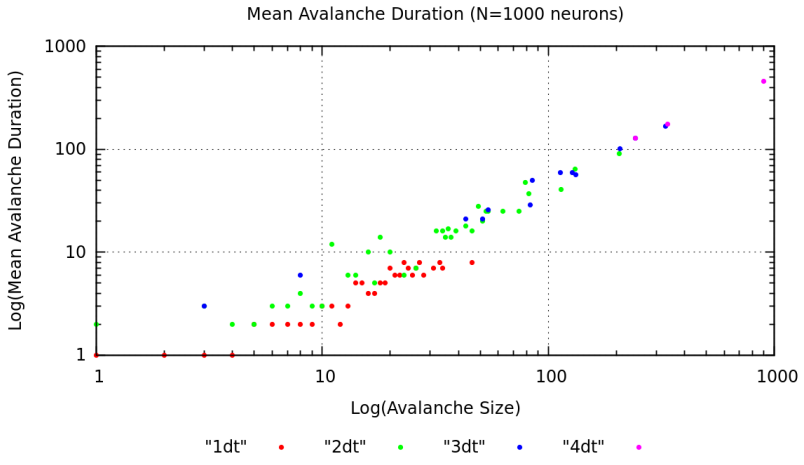


Figure 4.8: Mean avalanche duration for $1dt$, $2dt$, $3dt$ and $4dt$. $\gamma = 0.5$, $dt = 0.001$, 1000 neurons.

4.7.4 Avalanche Size

In this section we report the results for the probability $P_>(s)$ of having an avalanche size greater than s. The result is that for all the values of Δt the cumulative probability is an exponential.

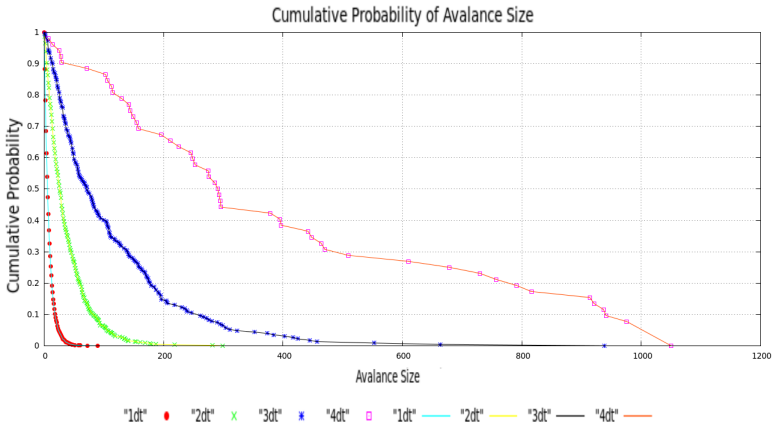


Figure 4.9: The cumulative probability of having an avalanche duration graeter than $t = ndt$ for different Δt of analysis, using a nonhomogeneous discrete Poisson process with $\gamma = 0.5$, $dt = 0.001$ and 1000 neurons.

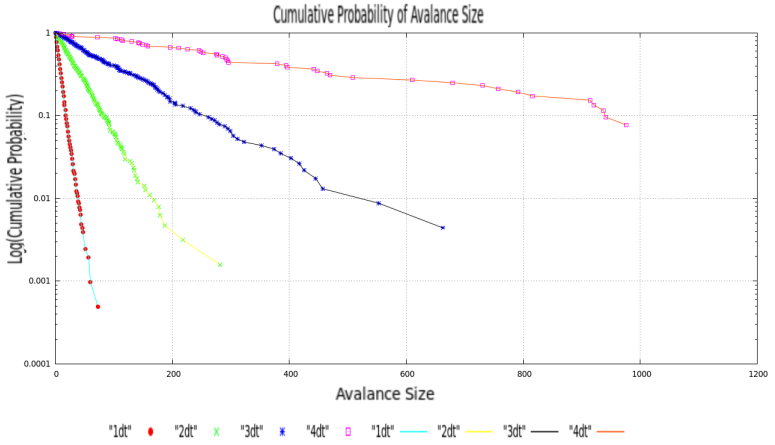


Figure 4.10: The same results of Fig.4.9 in Log-y-scale.

Note that varying Δt of analysis also the distribution varies, and it is obvious from Fig.4.9 that the behavior of the curve becomes less continuous increasing Δt.
In Table 5.2 there are all the informations about the various cumulative probability exponents for various Δt. The form of the distribution is, as for the avalanche duration, of the exponential type: $P_>(s) \simeq b^{-ks}$.

Δt	$1dt$	$2dt$	$3dt$	$4dt$	$5dt$
k	$-0.1 \pm 1\%$	$-3 \cdot 10^{-2} \pm 1\%$	$-9 \cdot 10^{-3} \pm 3\%$	$-2 \cdot 10^{-3} \pm 4\%$	$-5 \cdot 10^{-5} \pm 7\%$

Table 4.2: Exponents of the exponential distribution fit $P_>(s) \simeq e^{-ks}$ of avalanche sizes in the Poisson network reported in Fig.4.10.

4.8 Mean avalanche size

Now we want to know what is the mean size $\langle s \rangle$ of a neuronal avalanche with duration T. To do this, given the neuronal activity $N(t)$ we have to take all the avalanche with duration T and mediate their sizes. For example, if $N(t)$ admits k avalanches of duration T, then $\langle s \rangle = \frac{1}{k} \sum_{i=1}^{k} s_i$.
In this case the result is a power-law invariant under variations of Δt, as reported in Fig.4.11.

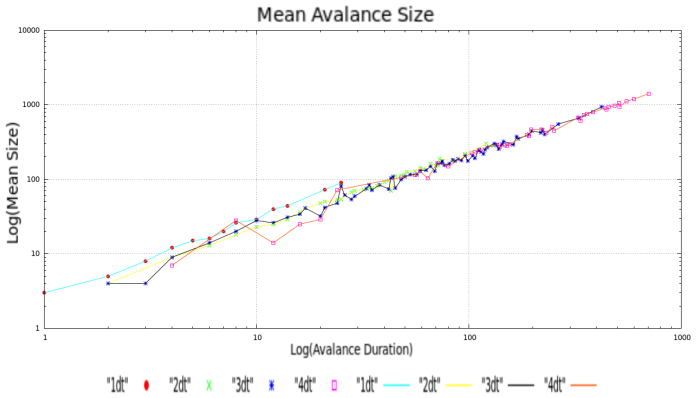


Figure 4.11: Mean avalanche size reported in Log-Log scale plot, note that different curves found with different Δt collapse in the same curve. $\gamma = 0.5$, $dt = 0.001$, 1000 neurons.

The exponent b of the interpolating distribution function $\langle s \rangle (T) = aT^b$ is for every distribution, varying Δt, $b = 1.08 \pm 0.01$. In this case it is evident that the power law distribution of $\langle s \rangle$ is almost as a linear relation between $\langle s \rangle$ and t and, moreover, the exponent b is the same for every distribution.
This is a curious case of scale invariance in absence of criticality in the system.

4.8.1 Average Shapes of the Avalanches

The last analysis that can be done in this model is to look inside the avalanches. How many neurons are activated at the first time step? How many at the second step?
The main goal of this section is to analyze the average number of neurons activated at every time step in a neuronal avalanche.

To do this, taken an avalanche of time duration $T = ndt$, we count how many neurons are activated in each of the n time steps, and average them. After this procedure, we obtained the average number of neurons active at every time step, we call it $\langle n(t_i, T) \rangle$ where $t_i = idt$, $i = 1, 2, 3 \ldots$, and we can observe the internal structure of every avalanche.

We can view from Fig.4.12 that the internal structure of the avalanche of duration $9dt$, $10dt$ and $11dt$ has not much informations because of the random structure of the model: every neuron is completely independent on the others and then the activation/deactivation

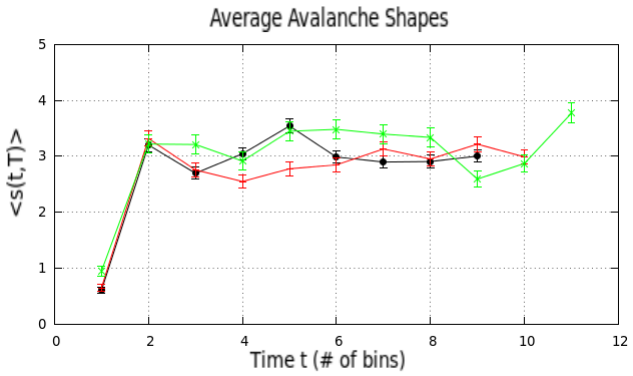


Figure 4.12: Average shapes for different time duration of neuronal avalanches: $9dt$ (black), $10dt$ (red) and $11dt$ (green). $\gamma = 0.5$, $dt = 0.001$, 1000 neurons.

of one neuron does not influence the activity of the other neurons
What is interesting is that all the curves in Fig.4.12 are collapsed in the same curve, as there is not difference between an avalanche of time duration T and one other of duration T'.
This is reasonable because the system is composed by N neurons completely random and independent each other, and there is not reasons for which one avalanche might be different from one other.
In the next chapter we will see how interactions between neurons modify this model, and how will be generated different types of statistics.

Chapter 5

The Ornstein-Uhlenbeck Neuronal Network

5.1 Markov Processes

"History is a cyclic poem written by time upon the memories of man."
Percy Bysshe Shelley

Markov processes provide a powerful lens for viewing the world. Provided a small number of assumptions (a fixed set of states, fixed transition probabilities, and the possibility of getting from any state to another through a series of transitions) a Markov process always converges to a unique distribution over states [**?**]. This means that what happens in the long run won't depend on where the process started or on what happened along the way. What happens in the long run will be completely determined by the transition probabilities – the likelihoods of moving between the various states.
If a system follows a Markov Process, then initial conditions, interventions, and history itself have no bearing on the long run distribution over states.

Definition 7. (Markov process). A stochastic process $X(t) \in \mathbb{R}$ is called Markov process if $\forall t_1 < t_2 < \cdots < t_n$ and $\forall n$, the conditional probability $p(x_n, t_n|x_1, t_1; x_2, t_2; \ldots; x_{n-1}, t_{n-1})$ satisfies the property:

$$p(x_n, t_n|x_1, t_1; x_2, t_2; \ldots; x_{n-1}, t_{n-1}) = p(x_n, t_n|x_{n-1}, t_{n-1}). \tag{5.1}$$

The above definition is equivalent to say that the event characterized by $\{x_n \leq X(t_n) \leq x_n + dx_n\}$ depends only on the previous event $\{X(t_{n-1}) = x_{n-1}\}$. In other words a Markov process does not depend on its whole history (stochastic process with no memory).

Lemma 5.1.1. *A Markovian stochastic process is completely determined by the one-point density probability $p(x, t)$ and by the conditional probability $p(x_2, t_2|x_1, t_1)$.*

Proof. We know that for a generic stochastic process the n-point joint distribution is given by

$$p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = p(x_1, t_1)p(x_2, t_2|x_1, t_1)p(x_3, t_3|x_2, t_2; x_1, t_1)\ldots \tag{5.2}$$

$$\ldots p(x_{n-1}, t_{n-1}|x_{n-2}, t_{n-2}; \ldots; x_1, t_1)p(x_n, t_n|x_{n-1}, t_{n-1}; \ldots; x_1, t_1).$$

By applying the Markov property to the previous equation we then obtain the lemma:

$$p(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) = p(x_1,t_1)p(x_2,t_2|x_1,t_1)p(x_3,t_3|x_2,t_2)\ldots \tag{5.3}$$

$$\ldots p(x_{n-1},t_{n-1}|x_{n-2},t_{n-2})p(x_n,t_n|x_{n-1},t_{n-1}).$$

□

Definition 8. A stochastic Markov process is stationary if

1. $p(x_2,t_2|x_1,t_1) = p(x_2,t_2-t_1|x_1)$.
2. $p(x,t) = p(x)$.

5.1.1 Processes with stationary and independent increments

Definition 9. (Process with independent increments). A stochastic process $X(t) \in \mathbb{R}$ has independent increments if, $\forall T = \{t_1,t_2,\ldots,t_n\}$

$$X(0), X(t_1)-X(0), X(t_2)-X(t_1), \ldots, X(t_n)-X(t_{n-1}) \tag{5.4}$$

are independent random variables.

Definition 10. (Process with stationary increments). A stochastic process $X(t) \in \mathbb{R}$ has stationary increments if $X(t)-X(s)$ has the same probability distribution than $X(t+\tau)-X(s+\tau)$, $\forall s,\ t,\ \tau \geq 0,\ s<t$.

If $X(t) \in \mathbb{R}$ is a stochastic process with stationary and independent increments the following properties hold

$$\mathbb{E}\{X(t)\} = \eta_1 t, \tag{5.5}$$

$$Var\{X(t)\} = \sigma_1 t \tag{5.6}$$

and

$$\mathbb{E}\{X(t)X(s)\} = \sigma_1 min(t,s) \tag{5.7}$$

where $\eta_1 = \mathbb{E}\{X(1)\},\ \sigma_1 = Var\{X(1)\}$.

5.2 The Ornstein-Uhlenbeck Process

Definition 11. (Gaussian process). A Gaussian, zero average, stochastic process $X(t) \in \mathbb{R}$ is a stochastic process in which for any given time partition $T = \{t_1,t_2,\ldots,t_n\}$ the joint probability density function is given by

$$p(x_1,t_1;\ldots,x_n,t_n) \equiv \frac{1}{(2\pi)^{n/2}}(detA)^{1/2}e^{-\frac{1}{2}\sum_{i,j=1}^{n} x_i A_{ij} x_i} \tag{5.8}$$

where $A \in Mn(\mathbb{R})$ is a real ($A_{ij} \in \mathbb{R}$), symmetric ($A_{ij} = A_{ji}$), invertible ($detA \neq 0$), strictly positive defined ($\sum_{i,j=1}^{n} x_i A_{ij} x_j > 0$), $n \times n$ matrix.

Definition 12. (Continuous in probability). A stochastic process $\{X(t)\}_{t\geq 0}$ is **continuous in probability** at t if $\forall \varepsilon > 0$,

$$\lim_{s\to t} \mathbb{P}\{\omega \in \Omega || X(s,\omega) - X(t,\omega)| \geq \varepsilon\} = 0. \tag{5.9}$$

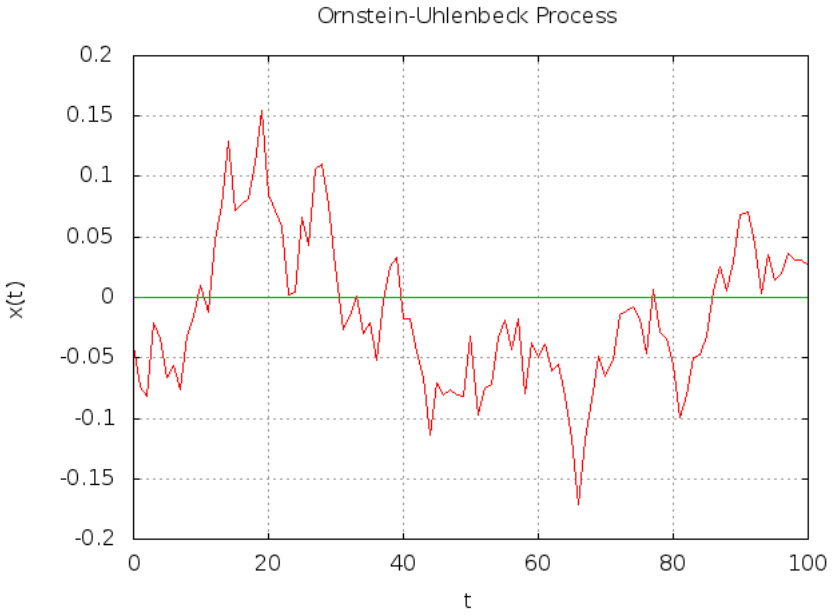


Figure 5.1: An example of Ornstein-Uhlenbeck process, with $\theta = 0,\ \sigma = 0.1,\ \theta = 1.5$.

The stochastic process Ornstein-Uhlenbeck is the most famous example of Markov process that admits a Gaussian distribution for $t \to \infty$ and that has in the limit a finite variance.

A stochastic process $\{X(t)\}_{t\geq 0}$ is an **Ornstein-Uhlenbeck** process or Gauss-Markov process if it is *stationary*, *Gaussian*, *Markovian* and *continuous in probability*. A fundamental theorem, due to Doob, ensures that $\{X(t)\}_{t\geq 0}$ necessarily satisfies the following linear stochastic differential equation (see Chapter on Langevin equation)

$$dX(t) = -\mu(X(t) - \theta)dt + \sigma dW(t) \tag{5.10}$$

where $dW(t) = W(t+\Delta t) - W(t)$ are the increments of a Wiener process $\{W(t)\}_{t\geq 0}$ with unit variance. An example of this type of process is reported in Fig. 5.2. The parameters $\mu,\ \theta$ and σ are constants. The moments of the Ornstein-Uhlenbeck process are

$$\mathbb{E}X(t) = \theta, \quad \mathbb{E}X(t)X(s) = \frac{\sigma^2}{2\mu} e^{-\mu|s-t|} \tag{5.11}$$

in the unconditional (stationary) case and

$$\mathbb{E}\{X(t)|X(0) = x_0\} = \theta + (x_0 - \mathbb{E}\{X(t)\})e^{-\mu t}$$

$$\mathbb{E}\{X(t)X(s)|X(0) = x_0\} = \frac{\sigma^2}{2\mu}\left(e^{-\mu|s-t|} - e^{-(s+t)}\right) \tag{5.12}$$

in the conditional (asymptotically stationary) case. Because of the form of $\mathbb{E}\{X(t)X(s)\}$, the Ornstein-Uhlenbeck process is also known as coloured noise.

5.3 The Neuronal network

Why is Ornstein-Uhlenbeck process useful for our analysis of neuronal networks? In the model that we will now develop, the firing rate $\lambda(t)$ of the neurons is given by the positive part $\lambda^+(t)$ of this process, while the negative part is fixed to zero.
The goal of this section is to reproduce the results of Johnatan Touboul and Alain Destexhe [4], using their same neuronal network model. We take the positive part of the following Ornstein-Uhlenbeck process:

$$d\lambda(t) = -\lambda(t)dt + dW(t) \tag{5.13}$$

and, for generating this process we pose $\alpha = \sigma = 1,\ \theta = 0$.
In Fig.5.2a we represent a typical pattern of the neuronal activity, every neuron spikes following a Ornstein-Uhlenbeck spiking rate. In this case, differently from the Poisson network, the Poisson process is continuous and then we have adopted the thinning algorithm described in the previous chapter for generating nonhomogeneous Poisson processes. So in this case,

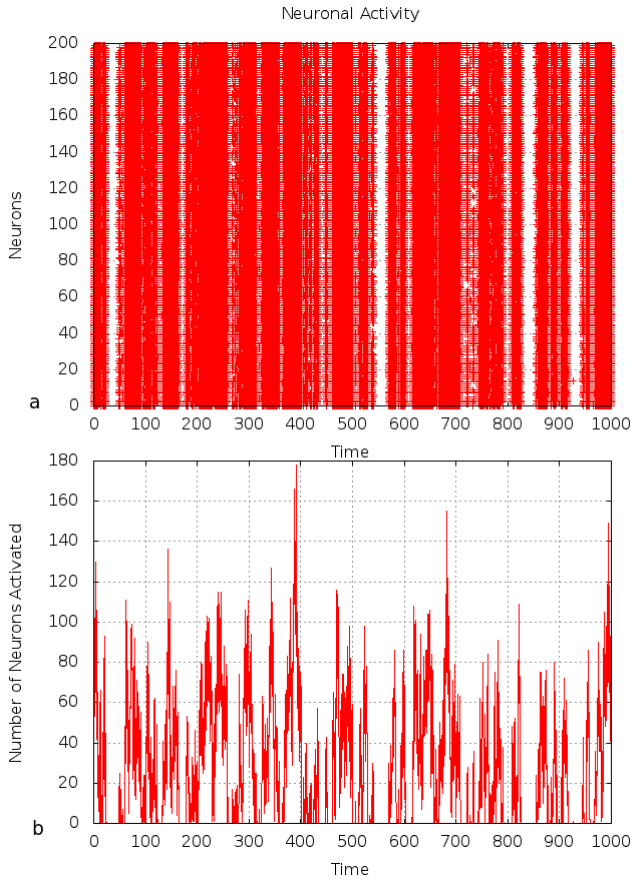


Figure 5.2: An example of Ornstein-Uhlenbeck neuronal activity for 200 neurons and $\mu = 1,\ \sigma = 1,\ \theta = 0.1$.

once generated $\lambda(t)$, we have calculated λ_{max} and for every neuron we have generated the following Poisson process

$$t_{i+1} = t_i - \frac{1}{\lambda_{max}} \ln(u) \tag{5.14}$$

at every time step $t_j \in [t_i; t_{i+1}]$ we evaluated the quantity $\lambda(t_j)$ by interpolation

$$\lambda(t_j) = \lambda_j = \lambda_i + \frac{\lambda_{i+1} - \lambda_i}{t_{i+1} - t_i}(t_j - t_i) \tag{5.15}$$

where, after discretized the time intervals, we have $t_{i+1} - t_i = dt$.
The choice of $\Delta t = ndt$ for the analysis of the avalanches is very important, as we will see computing the statistical distributions, and it must be fixed before doing the simulation.
In Fig.5.2b we can see the existence of the avalanches, in this case avalanches must exist because of the definition of the model: where $\lambda(t)$ is negative it is fixed to zero, in these intervals there will not exist neuronal activity.
As in the Poisson network, we define avalanche time as that time that elapses between the beginning of the neuronal activity and the ending of it. As neuronal activity is meant the quantity of neurons activated in a certain time interval.
Following this procedure, we have defined the probability $P_>(t)$ of having an avalanche duration greater than a time interval t. Note that this distribution is continue, but the time is discretized: in general, we can think that $t = ndt$, because of the discretization of the time through dt.

5.4 Avalanche Duration Statistics

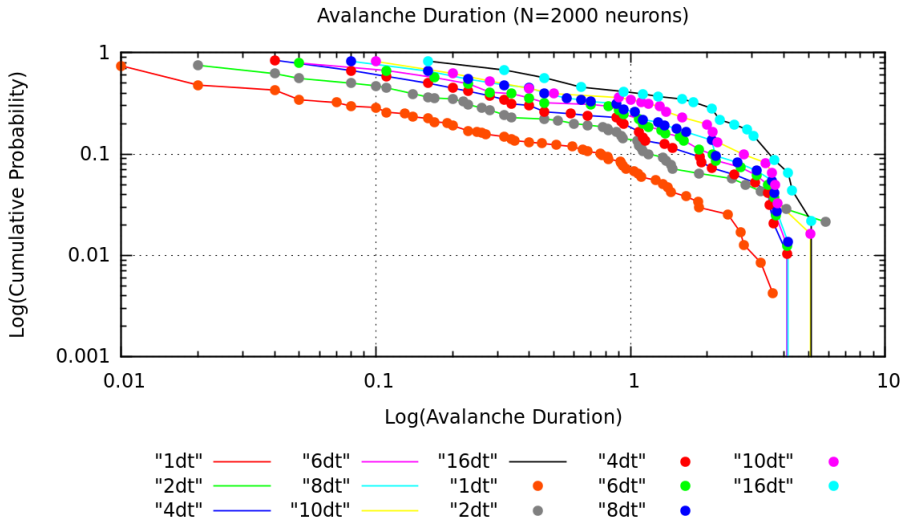


Figure 5.3: Avalanches duration statistics for different values n of time $\Delta t = ndt$ of analysis. $\mu = 1$, $\sigma = 1$, $\theta = 0$, $N = 2000$ neurons.

In Fig. 5.3 we found the cumulative probability $P_>(t)$ of the avalanche duration, we first see that is a power law distribution. Every distribution has different power law index, and it can be seen in the plot: changing the Δt of analysis the distribution changes its index.

This is the same result find by Johnatan Touboul and Alain Destexhe [4] that we already considered previously while we analyzed the probes against the hypothesis of criticality in living systems.

Δt	$1dt$	$2dt$	$4dt$	$6dt$	$8dt$	$10dt$	$16dt$
m	$-0.5 \pm 1\%$	$-0.4 \pm 3\%$	$-0.5 \pm 2\%$	$-0.5 \pm 6\%$	$-0.5 \pm 5\%$	$-0.4 \pm 5\%$	$-0.4 \pm 3\%$

Table 5.1: Exponents of the power law distribution fit $P_>(t) \sim t^m$ of avalanche duration of Fig.5.3.

We want to follow the same procedure, but searching a new contribute for a better comprehension. This result is a typical case in which a non critical system generates power law distributions. This system, as we will see, generates a plethora of power laws, although it has nothing that could lead it to criticality!

5.4.1 Analytical Calculations

Starting from the calculations done in the chapter of the Poisson network, we can define an analytical function which describes the avalanche duration probability.
In the case of the Ornstein-Uhlenbeck process there is a correlation between $Q(\lambda_i)$ and $Q(\lambda_{i-1})$ because of its Markovian nature: each step of the process depends only by the previous step, and so on. In this case $Q(\lambda_i|\lambda_{i-1}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\lambda_i - \lambda_{i-1})^2}{2\sigma^2}}$, and the probability takes the following form:

$$P_>(t) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \int \prod_{i=2}^{n} e^{-\frac{(x_i - x_{i-1})^2}{2\sigma^2}} \left[1 - e^{-\alpha x_i}\right] \prod_{j=1}^{n} dx_j \tag{5.16}$$

This is an integral that poses a lot of difficulties to be solved, because of the correlation between λ_is typical of Markov processes. If it could be solved, it would give us the analytical probability function to be compared with the computational solution. However, until today we have not found a method to solve it.

The particular case $\lambda = const$

It is interesting to verify, in a very particular case, that the correlations are the key ingredient for obtaining power laws: this is the case of completely correlated λs, i.e. when $\lambda = const$.
The equation

$$P_>(n) = \int \prod_{i=1}^{n} dQ(\lambda_i)[1 - (1 - \lambda_i dt)^N], \tag{5.17}$$

becomes for $\lambda = const$

$$P_>(n) = \int d\lambda (1 - e^{-\alpha\lambda})^n Q(\lambda) \tag{5.18}$$

Now, we can consider exponential $Q(\lambda)$:

$$Q(\lambda) = \gamma e^{-\gamma\lambda} \tag{5.19}$$

For $\lambda = const$ we have

$$P_>(n) = \gamma \int_0^\infty d\lambda e^{-\gamma\lambda} (1 - e^{-\alpha\lambda})^n \tag{5.20}$$

Where $\alpha = N\delta t$,

$$= \gamma \int_0^\infty d\lambda \exp(-\gamma\lambda + n \ln(1 - e^{-\alpha\lambda})) \tag{5.21}$$

The integral is of the following type:

$$P_{>}(n) = \int_0^{\infty} e^{-f(x)}\, dx \tag{5.22}$$

The integral is dominated by the minimum of the function $f(x)$:

$$\lambda_{min} = \frac{1}{\alpha} \ln\left(1 + \frac{\alpha n}{\gamma}\right) \tag{5.23}$$

So we have

$$P(n) \sim \exp\left(-\gamma\lambda_{min} + n \ln\left(1 - e^{-\alpha\lambda_{min}}\right)\right) = \tag{5.24}$$

$$= \exp\left(-\frac{\gamma}{\alpha} \ln\left(1 + \frac{\alpha n}{\gamma}\right) + n \ln(1 - \left(1 + \frac{\alpha n}{\gamma}\right)^{-1})\right) \tag{5.25}$$

Where $\ln\left(1 - \left(1 + \frac{\alpha n}{\gamma}\right)^{-1}\right) \simeq -\frac{\gamma}{\alpha n}$. The result is

$$P(n) \simeq \left(1 + \frac{\alpha n}{\gamma}\right)^{-\gamma/\alpha} \sim_{n \gg \frac{\gamma}{\alpha}} n^{-\gamma/\alpha} = n^{-\gamma/Ndt} \tag{5.26}$$

The result is that in this particular case we obtain a power law. For this reason we are tempted to think that the difference from a completely random Poisson network in the Ornstein-Uhlenbeck rate are the correlations, and they are the indispensable to have a power law relation statistics.

5.4.2 Mean Avalanche Duration

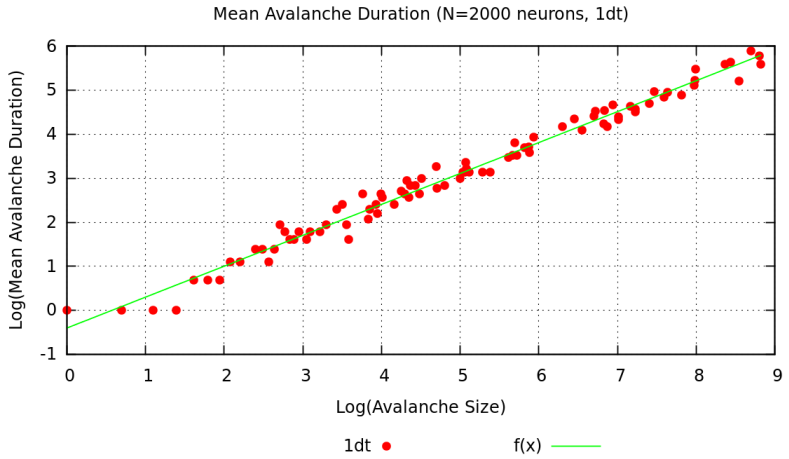


Figure 5.4: Mean avalanche duration for $\Delta t = 1dt$. $\mu = 1$, $\sigma = 1$, $\theta = 0$, $N = 2000$ neurons. In green it can be seen the fit line.

It is interesting to find what is the behavior of the mean duration of an avalanche given a certain size. To find this relation, as for the Poisson neuronal model we calculated the mean

Δt	$1dt$	$2dt$	$4dt$	$6dt$	$8dt$	$10dt$	$16dt$
m	$-0.6 \pm 4\%$	$-0.5 \pm 5\%$	$-0.3 \pm 8\%$	$-0.3 \pm 6\%$	$-0.3 \pm 7\%$	$-0.2 \pm 13\%$	$-0.2 \pm 25\%$

Table 5.2: Exponents of the power law distribution fit $P_>(s) \sim s^m$ of avalanche size of Fig.5.6.

duration $\langle t \rangle$ of an avalanche, for a given s:

$$\langle t \rangle (s) = \frac{1}{n} \sum_{i=1}^{n} t_i(s) \tag{5.27}$$

In Fig.5.4 it is reported the case for $\Delta t = 1dt$, we can see the linear relationship in the logarithm plot: it is a good power law relation $\langle t \rangle (s) \simeq s^m$ with $m = 0.70 \pm 0.02$ given by the fit of the data.

In Fig.5.5 we can view how the mean avalanche duration changes with the changing of Δt, it is fascinating to observe how the data asymptotically collapse on the straight line given in Fig.5.4, as they all tend to have this type of behavior for big avalanche sizes.

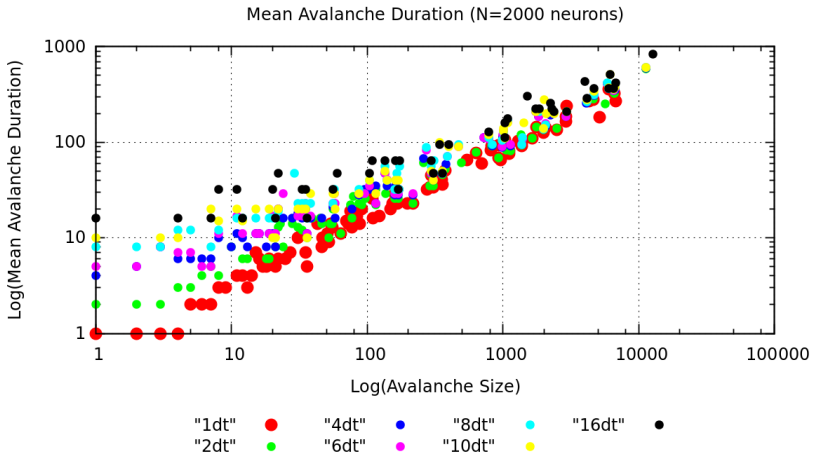


Figure 5.5: Mean avalanche duration for different values of Δt. $\mu = 1$, $\sigma = 1$, $\theta = 0$, $N = 2000$ neurons. In red it can be seen the fit line.

5.5 Avalanche Size Probability

We want to view what is the behavior or the probability $P_>(s)$ of having avalanche sizes bigger than s.
To obtain this probability we calculate how many avalanches have size smaller than a fixed s, we calculate $P_<(s)$ an then we take the complementary probability: $P_>(s) + P_<(s) = 1$.
In Fig.5.6 it is reported $P_>(s)$, we can view the strong power law behavior with finite size scaling of the system.

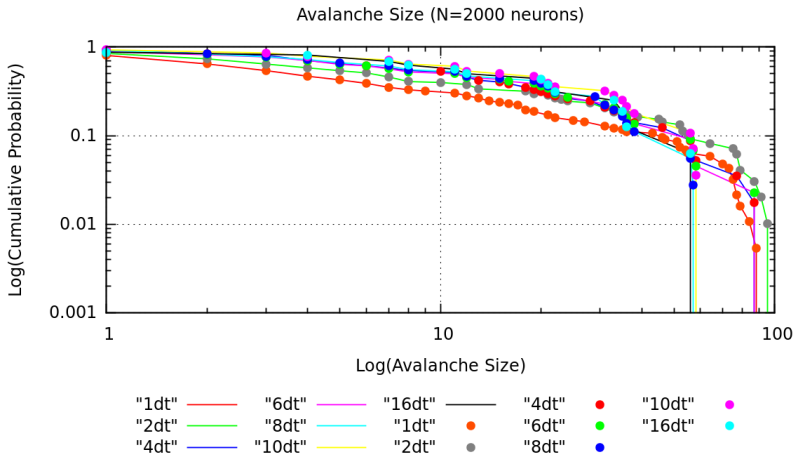


Figure 5.6: Avalanche size statistics for different values n of time $\Delta t = ndt$ of analysis. $\mu = 1,\ \sigma = 1,\ \theta = 0,\ N = 2000$ neurons.

5.6 Mean Avalanche Size

As we have done for the avalanche mean duration, we calculate the mean avalanche size $\langle s \rangle$ for a given duration t:

$$\langle s \rangle (t) = \frac{1}{n} \sum_{i=1}^{n} s_i(t) \tag{5.28}$$

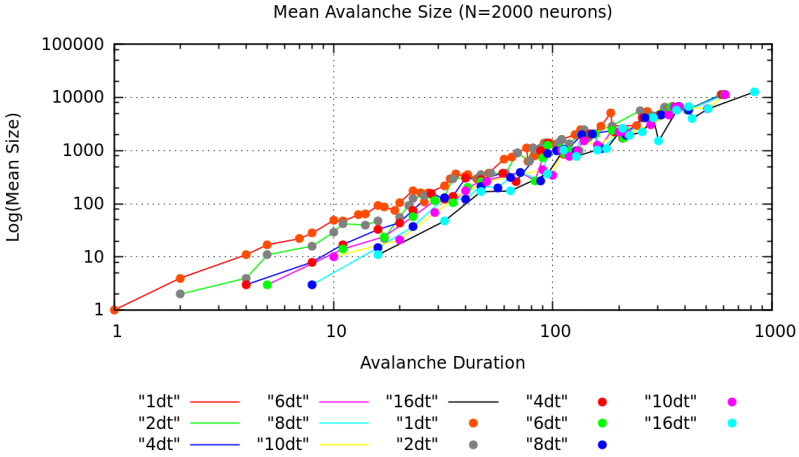


Figure 5.7: Mean Avalanche size statistics for different values n of time $\Delta t = ndt$ of analysis. $\mu = 1,\ \sigma = 1,\ \theta = 0,\ N = 2000$ neurons.

Also in this case we obtained a power law relation $\langle s \rangle (t) \simeq t^m$, all the curves have exponent $m = 1.48 \pm 0.02$, it is as the distribution behavior does not depend on the choice of Δt.

5.7 Average Shapes of the Avalanches

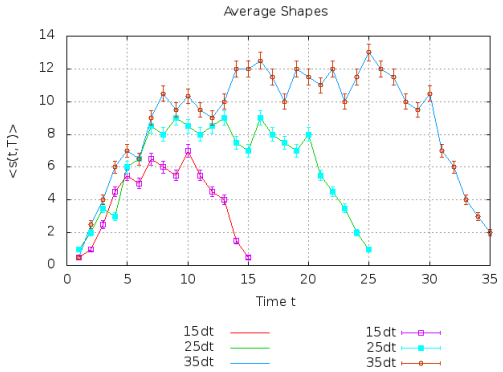


Figure 5.8: Average neuronal activity at each time step for avalanches of duration $15dt$, $25dt$ and $35dt$.

As we have done for the exponential distribution, we want now to view the internal structure of the avalanches, in particular for avalanches of duration $15dt$, $25dt$ and $35dt$. In Fig.5.8 are described the avalanches of time durations T at each time step $t = ndt$, it can be seen that every avalanche has the form of a "cupola", or rather in the first time steps there is an incrementing mean neuronal activity $\langle s(t,T)\rangle$ that reaches its maximum in the center of the avalanche and then decreases in the final steps.
This mirrors the behavior of the rate $\lambda(t)$ generated by the Ornstein-Uhlenbeck process. The Ornstein-Uhlenbeck rate $\lambda(t)$ is the reason for what the average shapes are different from the exponential distribution's case, in which every time step in the avalanche is completely random and uncorrelated from the previous.

After having obtained the average shapes of some avalanches, we want no to collapse one other the other. To do this, we report in Fig.5.9 the average neuronal activity for every time step rescaled with $T^{\gamma-1}$ as a function of the rescaled time steps t/T and it is found $\gamma = 1.6$.

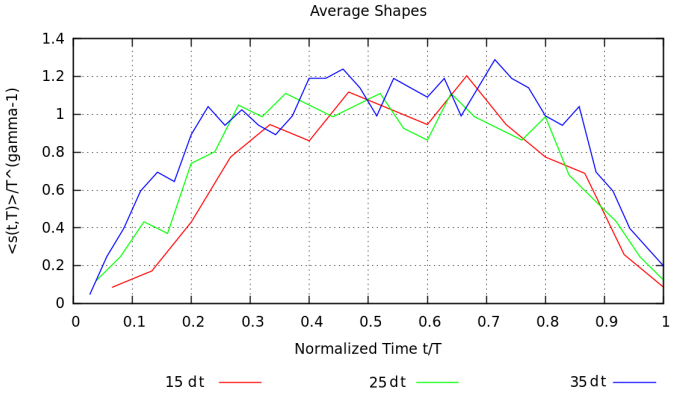


Figure 5.9: The collapse of the curves in Fig.5.8 with $\gamma = 1.6$.

5.8 Entropy and Energy

Following the method illustrated by Thierry Mora [53], now we search a signature of criticality in this neuronal model.
To do this, we used the pattern in Fig.5.2A for different numbers of neurons N searching for a relationship between the entropy and the energy of the neuronal system. We defined a neuronal state $\bar{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$, $\sigma_i = \{0, +1\}$, where $\sigma_i = 0$ if the neuron is switched off and $\sigma_i = +1$ if it is switched on. We do this at every time step dt, in particular every component σ_i tells about the neuron's activity in the interval $[t, t + dt]$.
First, we calculated the most frequent state $\bar{\sigma}_0$ in the system and, once calculated its probability p_0, we assigned to every state a certain energy that depends on its probability:

$$E(\bar{\sigma}) = -\ln\left(\frac{p_i}{p_0}\right) \tag{5.29}$$

where p_i is the probability to find the state $\bar{\sigma}_i$. The correspondent entropy of a state with

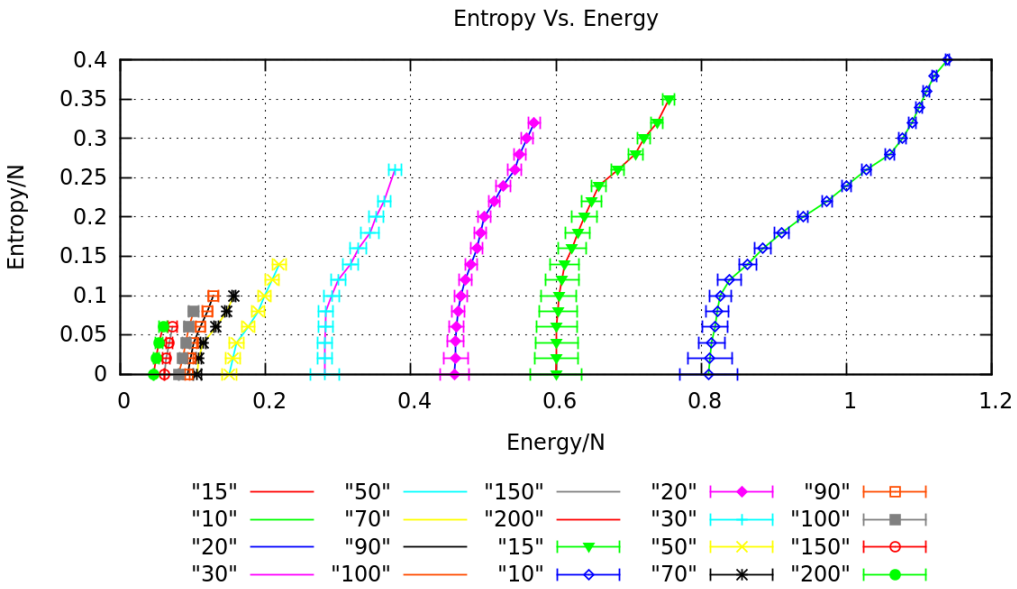


Figure 5.10: Entropy for neuron in function of the energy for neuron for different numbers of neurons.

energy E has been evaluated calculating the number of states $\mathcal{N}(E)$ that have energy less than E

$$S(E(\bar{\sigma})) = \ln \mathcal{N}(E(\bar{\sigma})) \tag{5.30}$$

where

$$\mathcal{N}(E(\bar{\sigma})) = \sum_s \theta(E(\bar{\sigma}) - E_s) \tag{5.31}$$

and $\theta(x)$ is the Heaveside theta function

$$\theta(x) = \begin{cases} 1 & x > 0 \\ \\ 0 & x < 0 \end{cases} \tag{5.32}$$

In most systems, including the networks that we study here, there are relatively few states that have high probability, and many more states with low probability; mathematically, $\mathcal{N}(E)$ is an increasing function. At large number of neurons N , this competition between decreasing probability and increasing numerosity picks out a special value of $E = E^*$, which is the energy of the "typical" states that we actually see; E^* is the solution to the equation

$$\frac{dS(E)}{dE} = 1 \tag{5.33}$$

For most systems, the energy $E(\sigma)$ has only small fluctuations around E^* and in this sense most of the states that we see have the same value of log probability per degree of freedom. But hidden in the function $S(E)$ are all the parameters describing the interactions among the N degrees of freedom in the system. At special values of these parameters, $[d^2S(E)/dE^2]_{E=E^*} \to 0$, and the variance of E diverges as N becomes large. This is a critical point, and is mathematically equivalent to the divergence of the specific heat in an equilibrium system.
In Fig.5.10 we can view the behavior of the entropy for neuron S/N as a function of the energy for neuron E/N. In Fig.5.11 is reported the comparison between the results of Thierry Mora [53] and the results of this neuronal model. We can appreciate that, although varying the number of neurons, $S(E)/N$ has not the typical asymptotic trend to $dS(E)/dE = 1$. Unfortunately due to the lack of additional data we can not appreciate the total behavior. We can deduce from the Fig.5.11 and calculate the slope of $S(E)$, and assume that this system as reported in Fig.5.10 does not exhibit criticality. It can be seen because the slope of $S(E)$, when the curves approach to the asymptotic behavior, is never unitary and has a minimum value of 1.70 ± 0.01 for 10 neurons. So, this is very different from that which provides the theory for a critical system.
For a better comprehension of the model, we would need much more data to view the complete behavior of the system. To do this, there would be some useful models that can be applied to get much more data from this neuronal model.

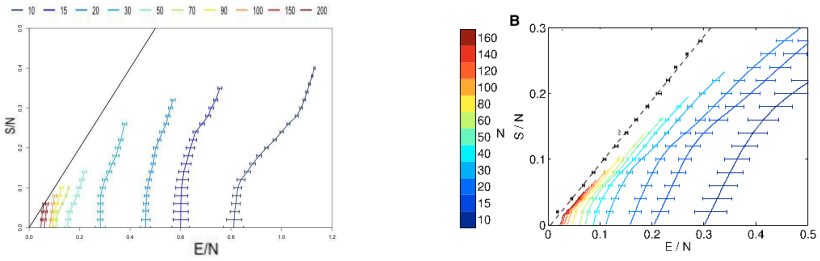


Figure 5.11: Comparison between the results for $S(E)$ for the neuronal model with Ornstein-Uhlenbeck rate $\lambda(t)$ and the results of Thierry Mora [53].
Fig.**A**: Computed from the data of the Ornstein-Uhlenbeck neuronal model, with varying number of neurons: 10, 15, 20, 30, 50, 70, 90, 100, 150 and 200 neurons. The minimum slope of the curves is equivalent to 1.70 ± 0.01.
Fig.**B**: Computed directly from the data used by Thierry Mora. Different colors show results for different numbers of neurons; in each case are chosen 1000 groups of size N at random, and points are means with standard deviations over groups. Inset shows extrapolations of the energy per neuron at fixed energy per neuron, summarized as black points with error bars in the main figure. Dashed line is the best linear fit to the extrapolated points, $S/N = (0.974 \pm 0.02)(E/N) + (-0.005 \pm 0.003)$.[53]

With this thermodynamic analysis we have demonstrated that this neuronal model is not posed at a critical state, in spite of everything it exhibits all power law statistics. With this

example we have viewed that a simple neuronal model, as the model with rate generated by a Ornstein-Uhlenbeck process, can show power law statistics in a lot of its facets but at the same time it is not posed at a critical state. For this reason, it is very important to know something more about the system, because some living systems could show power laws although they are not poised at a critical state.

Chapter 6

Conclusions

Criticality is a particularly relevant phenomenon in the context of statistical mechanics, with its characteristic power law distributions and free scale behavior. For some living systems, it seems that the critical state is the best state to be placed on, and scientists have hypothesized that being poised at such special state optimizes several tasks [52, 2]. On the other hand, it has been proved that some non critical systems exhibit power law distributions [4]. In this scenario, the question has been place: "are living systems poised at a critical state?"

Living organisms constitute an amazingly complex physical system, where internal interactions play a fundamental role.

In the context of neural networks, we have shown that a system with Poisson-like neurons with rate generated from an exponential distribution does not exhibit power law statistics, because of the absence of interactions between its units. On the other hand, introducing such interactions immediately generates power law statistics. Such power law distributions are very special from the point of view of statistical mechanics, being the characteristic fingerprint of a system poised at a critical point. From this perspective, it seems that systems with a particular interaction pattern may be poised at a critical state, with the consequent emergence of power law statistics.

However we have seen that, implementing the framework of Mora et al. [53] in our neuronal model with a firing rate generated from a special Markov process, although many power law statistics are present, the system is not critical at all. In fact, the model used has nothing of critical in its structure.

We have seen that being critical implies the emergence of power laws, but the opposite is not true. In fact, it is known that many non-critical systems exhibit power-law behavior [4]. It could be useful to analyze these power-law generating mechanisms under the framework of "maximum entropy models" (see Appendix 2) that can infer the values of the fundamental parameters behind the dynamics. These models could be useful to understand if the parameters related to that models are typical of a critical system or not, associating these parameters to an Ising model and studying the system in this way.

We could be tempted to think that living systems are poised at a critical state because of the existence of a lot of power laws in the models that describe them. What we have found is that the relation ($criticality \leftrightarrow Power\ laws$) is not bijective and some particular systems can exhibit power laws even if they are not poised at a critical state.

Furthermore, it is possible to see how the choice of the temporal window in the data analysis is so influential in the values of the exponents of the distribution, that change significantly with the variation of such binning, thus weakening the possibility of identifying universality classes in the dynamics.

The possible question for the future is: "*how can we distinguish a critical system from a non critical system from its parameters?*" To have an answer we need a more precise analysis of these systems, and to do this we need further effective models to understand better the type of interactions that characterizes a critical system. Not all the living systems are poised at a critical state, but there might exist a parameter that regulates and characterizes them when they are critical.

Chapter 7

Thanks to...

In this section I would like to thank all the people that have supported me during these years and for simplicity's sake I will now switch to my mother language.

Il primo ringraziamento va a tutto il gruppo di ricerca che mi ha accolto in un modo unico e mi ha fatto crescere moltissimo.

Un grazie al Professor Maritan, per la sua costante disponibilità e per avermi sostenuto anche nei momenti più importanti.

Grazie a Samir, per le intere mattinate che ha dedicato alla discussione dei risultati che ottenevo e per i suoi preziosi consigli.

Grazie a Jordi, per i suoi indispensabili consigli sulla programmazione.

Un grazie ancora a tutto il gruppo di ricerca, per aver reso questo periodo di tesi così appassionante e pieno di occasioni per apprezzare sempre di più questo ambito della Fisica.

Un immenso grazie alla mia famiglia, per avermi sempre sostenuto e per i sacrifici fatti per farmi raggiungere questa importante tappa della mia vita.

Un grazie alla Professoressa Nardin, che mi segue da sempre e mi ha trasmesso la passione per la Fisica.

Grazie a tutti i miei amici, a quelli con cui ho condiviso ogni minima parte di questo fantastico percorso, a quelli con cui ho passato le notti a studiare e a cercare gli errori nelle relazioni di laboratorio, a quelli con cui ho condiviso le delusioni e i successi, a quelli che hanno sempre avuto un momento per ascoltarmi e darmi un consiglio che mi ha fatto crescere, agli insostituibili compagni di collegio, ai miei amici arbitri, a quelli che hanno dato prova che a volte l'amicizia segue un andamento a power law (con esponente positivo) della distanza relativa ed è una funzione indipendente dal tempo, e a quelli che nonostante tutto ci sono sempre stati.

Un grazie anche a tutte quelle persone che hanno avuto un ruolo importante nella mia vita.

"Ogni persona che passa nella nostra vita è unica. Sempre lascia un pò di sè e si porta un pò di noi. Ci sarà chi si è portavo via molto, ma non ci sarà mai chi non avrà lasciato nulla. Questa è la più grande responsabilità della nostra vita e la prova evidente che due anime non si incontrano mai per caso."(J. Luis Borges)

Appendix A

Appendix 1: Attractors in Randomly connected networks

Neurons in the brain interact with each other in a heterogeneous and asymmetric way producing complex neuronal activity. The dynamics is given by the equations of the network, and it exhibits fixed-point behavior, limit cycles or high-dimensional chaos.
In these networks, connections between binary neurons are independently drawn from an identical distribution, and the state of a network is updated simultaneously in discrete time steps without thermal noise. Every initial configuration must evolve into an attractor, which is a fixed point or a limit cycle. The typical length of the cycles was observed to grow exponentially with the number of neurons n (such kinds of cycles are called chaotic attractors), and the total number of attractors increases linearly with n.
These quantities were also analytically evaluated based on an empirical assumption that the dynamics of the system loses memory of its nonimmediate past. Now we are gonna characterize the attractors through a mean-field theory of the asymmetric neuronal network by extending the state concentration concept.

A.1 The Model

Consider randomly connected neuronal networks of n neurons (units). Each unit interacts with all the other units with an asymmetric coupling: we use J_{ij} to represent the coupling strength from unit j to i, it is independent of J_{ji} and they follow the same Gaussian distribution $N(0, \frac{1}{n})$.
The state of neuron i $(i = 1, \ldots, n)$ at time $t + 1$ is determined by the following equation:

$$\sigma_i(t+1) = sgn\left[h_i(t)\right] = sgn\left[\sum_{j=1}^{n} J_{ij}\sigma_j(t)\right] = \begin{cases} +1 & (activate\ state) \\ \\ -1 & (silent\ state) \end{cases}$$

In a mean-field analysis, consider now the dynamical evolution of the overlap between two states along a trajectory, expecting that its distribution across different realizations of J_{ij} contains information about the structure of attractors. Let us define the overlap of two states, say $\bar{\sigma}(t) = \sigma_i(t)/i = 1, \ldots, n$ and $\bar{\sigma}(s) = \sigma_i(s)/i = 1, \ldots, n$ along the same trajectory at different times $t > s$,

$$q_{ts} \equiv \frac{1}{n}\sum_{i=1}^{n} \sigma_i(t)\sigma_i(s) \tag{A.1}$$

The overlap takes the values ± 1 respectively if two states are the same or one is the sign flip of the other. The overlap takes discrete values in finite-size networks, but can be approximated as a continuous quantity in the large network size limit. The dynamics of this quantity is described by the mean-field theory, consisting of the dynamics of the overlap parameter and its fluctuations defined over the ensemble of random J_{ij}.
The stochastic dynamics of the overlap is well approximated for large n by a Markovian process

$$P_{t+1,s+1}(q) \simeq \int W(q|q')P_{ts}(q')dq'$$

where $P_{ts}(q) \equiv Prob(q_{ts} = q)$. The transition probability is approximated for large n by a simple binomial distribution

$$W(q|q') = \binom{n}{n(1+q)/2} \left[\frac{1+\varphi(q')}{2}\right]^{n(1+q)/2} \left[\frac{1-\varphi(q')}{2}\right]^{n(1-q)/2} \simeq \tag{A.2}$$

$$\simeq \exp\left[n\left(H(q) + \frac{1+q}{2}\ln\frac{1+\varphi(q')}{2} + \frac{1-q}{2}\ln\frac{1-\varphi(q')}{2}\right)\right],$$

where $\varphi(q) = (2/\pi)\arcsin q$ and $H(q) \equiv -\frac{1+q}{2}\ln\frac{1+q}{2} - \frac{1-q}{2}\ln\frac{1-q}{2}$.
Note that this equation summarizes the probability that $n(1+q)/2$ out of n neurons take the same sign in states $\sigma(t+1)$ and $\sigma(s+1)$, given that $n(1+q')/2$ out of n neurons take the same sign in the previous steps. The binomial distribution suggests that the state overlap for each neuron is approximately independent, occurring with probability $\left[\frac{1+\varphi(q')}{2}\right]$.
The dynamics of the overlap becomes deterministic in the limit of large n, according to Central Limit Theorem.
The previous Markovian process provides sequentially $P_{t+l,t}(q)$ for $t = 1, 2, \ldots$ for some positive time difference $l = t - s$ given an initial distribution at $t = 0$, $P_l, 0(q) = Prob(q_{l,0} = q)$. Since the initial state $\bar{\sigma}(0)$ is randomly selected and independently from J_{ij}, we can set $\sigma_i = 1,\ \forall i$. In this case,

$$q_{l,0} = \frac{1}{n}\sum_{i=1}^{n}\sigma_i(l)\sigma_i(0) = \frac{1}{n}\sum_{i=1}^{n} sgn[h_i(l-1)]$$

If l is small, $P_{l,0}(q)$ reflects the memory of the initial state and is hard to evaluate exactly. However, if l is large, the mean-field approximation indicates that $h_i(l-1)/i = 1, 2, \ldots, n$ follows approximately a zero-centered Gaussian distribution with unit variance. $P_{l,0}(q)$ tends for large l to a binomial distribution $\binom{n}{n(1+q)/2}2^{-n}$, where the probability of $q_{l,0} = \pm 1$ is approximately 2^{-n} in the large-network limit, according to numerical simulations.
Now we demonstrate why the choice of the initial state $\bar{\sigma}(o)$ is completely arbitrary, and it is not important to study the dynamics of the state overlap. Without losing generality, we can set $\sigma_i = 1,\ \forall i$. Consider then the following transformation:

$$\tilde{\sigma}_i(t) = \sigma_i(t)\sigma_i(0)$$

The state overlap is also described in terms of these variables, and the transformed variables follow the same update rule as the original one,

$$\tilde{\sigma}_i(t+1) = sgn\left(\sum_j \tilde{J}_{ij}\tilde{\sigma}_j(t)\right)$$

except that the coupling matrix is given by $\tilde{J}_{ij} = \sigma(0)J_{ij}\sigma_j(0)$ instead of J_{ij}. The distribution is the same as that of J_{ij} as long as $\bar{\sigma}(0)$ is chosen independently of J_{ij}. Then, to study

the dynamics of the state overlap, we can study the dynamics of the transformed variables with the condition $\tilde{\sigma}_i(0) = 1/1 = 1, 2, \ldots, n$.

Now consider how different states concentrate in time, so taking in account the Markovian dynamics it is important to note that it is completely characterized by the eigenvalues and the eigenvectors of $W(q|q')$. Ranking eigenvalues in descending order, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n+1}$ (the number of possible values for q is $n+1$), the distribution of the overlap is expressed by a weighted sum of the eigenvectors as

$$P_{t+l,t}(q) = \sum_{a=1}^{n+1} (\lambda_a)^t A_a f_a(q) \tag{A.3}$$

where A_a is a set of initial coefficients that satisfies $P_{l,0}(q) = \sum_a A_a f_a(q)$. As time increases, $P_{t+l,t}(q)$ becomes progressively dominated by the components with large eigenvalues.
It is trivial to find that W has two eigenvectors of the form $f_1(q) = \delta_{q,1}$ and $f_2(q) = \delta_{q,-1}$ with degenerate eigenvalues $\lambda_1 = \lambda_2 = 1$. The third eigenvector $f_3(q)$ is a nontrivial one and its eigenvalue $\lambda_3 \simeq 1 - \exp(-0.41n)$ exponentially approaches to 1 with n. The fourth eigenvalue converges to $\lambda_4 \simeq 0.67$ in the limit of large n. The half-decay time of the a-th component is given by $(\lambda_a)^{t_a} = 1/2$.
The difference between the third and the fourth component indicates that, for large n, the distribution of the overlap must approach to the quasistationary state $P_*(q) \equiv \sum_{a=1}^{3} A_a f_a(q)$ at around $t_4 \simeq 1.73$ and stay unchanged until $t_3 \simeq 0.69 \exp(0.41n)$. In particularly, the quasistationary state is characterized solely by $f_3(q)$ except at $q = \pm 1$. When the mean-field theory breaks down, the theory is not applicable once the third eigencomponent significantly decays around t_3.

We want now to characterize more in detail the quasistationary state in large n limit, from which we extract the structure of the attractors. Introducing the auxiliary notation

$$\alpha_{t+l,t}(q) \equiv \frac{1}{n} \ln P_{t+l,t}(q),$$

where $\int \exp[n\alpha_{t+l,t}(q)]dq = 1$, according to normalization constraint. With this notation, we can express the dynamics of the Markovian process through the following equation:

$$\alpha_{t+l+1,t+1}(q) = \frac{1}{n} \ln \int W(q|q') P_{t+l,t}(q') dq' \simeq \tag{A.4}$$

$$\simeq H(q) + \max_{q'} \left[\frac{1+q}{2} \ln \frac{1+\varphi(q')}{2} + \frac{1-q}{2} \ln \frac{1-\varphi(q')}{2} + \alpha_{t+l,t}(q') \right],$$

where, assumed large n, in the second line has been applied the Laplace Method. The well-defined asymptotic solution is

$$\alpha(q) = H(q) + \max_{q'} \left[\frac{1+q}{2} \ln \frac{1+\varphi(q')}{2} + \frac{1-q}{2} \ln \frac{1-\varphi(q')}{2} + \alpha(q') \right],$$

with finite $\alpha(q)$ self-consistently provides the quasistationary state. The last equation permits an arbitrary discontinuity at $q = \pm 1$, reflecting that these values are the sink of the Markovian process.
Next we define the probability that two states $\bar{\sigma}(t+l)$ and $\bar{\sigma}(t)$ have overlap $q_{t+l,t}$ before converging in the next step ($q_{t+l+1,t+1} = 1$) as $\beta_{t+l,t}(q_{t+l,t}) \equiv \frac{1}{n} \ln Prob(q_{t+l,t}|q_{t+l+1,t+1} = 1)$, this index is expressed using the Bayes theorem as

$$\beta_{t+l,t}(q') = \frac{1}{n} \ln \frac{W(1|q') P_{t+l,t}(q')}{P_{t+l+1,t+1}(1)} = \ln \frac{1+\varphi(q')}{2} + \alpha_{t+l,t}(q') - \alpha_{t+l+1,t+1}(1) \tag{A.5}$$

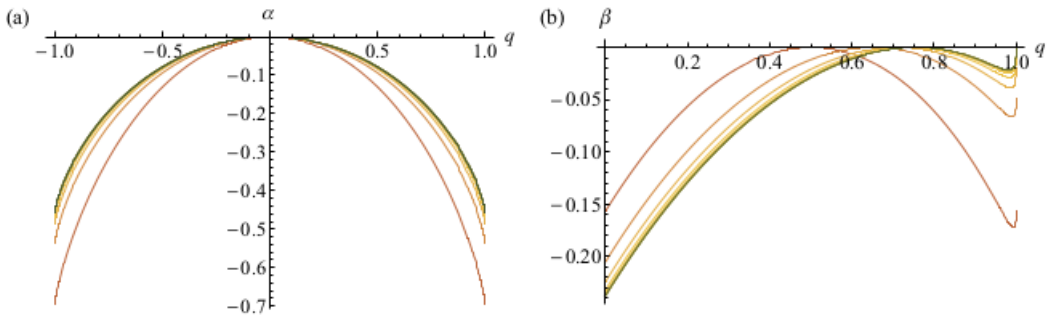


Figure A.1: (Color online) The Markovian dynamics of $\alpha_{t+l,t}$ and $\beta_{t+l,t}$ in time. (a) The index $\alpha_{t+l,t}$ characterizes the dynamics of distribution $P(q_{t+l,t})$. The line color changes from the lowest curve (the orange curve at $t = 0$, i.e., $\alpha_{l,0}$ or $\beta_{l,0}$) to the yellow, and finally to the gray (the top curve at $t = 10$, i.e., $\alpha_{l+10,10}$ or $\beta_{l+10,10}$). (b) The index $\beta_{t+l,t}$ characterizes the dynamics of distribution $P(q_{t+l,t}|q_{t+l+1,t+1} = 1)$. The result indicates that states concentrate mainly from $q \simeq 0.5$ at the beginning but concentrate equally from $q \simeq 0.75$ and $q \simeq 1$at the quasistationary state. We used $\alpha_{l,0}(q) = H(q) - \ln 2$ as the initial condition assuming no correlations at starting points. The results hold for any $l \geq 1$. [6]

This means that, for large values of n, most of the trajectories that lead to state concentration had an overlap specified by the peak location of $\beta_{t+l,t}(q)$; in the case of randomly connected neuronal networks it has two peaks, one at $q = 1$ and the other at $q < 1(\simeq 0.75)$ and become comparable around t_4. These dynamics of the state overlap reflects the specific structure of attractors.

A.2 Statistical Properties of attractors

In this section we are gonna analyze the statistical properties of the attractors for randomly connected neuronal networks, using the state concentration probability. $p_{t+1,s+1}$ also describes the probability of $\bar{\sigma}(t+1) = -\bar{\sigma}(s+1)$. Hence,

$$p_{t+1,s+1} \equiv Prob(q_{t+1,s+1} = \pm 1 | \{q_{t',s'} \neq \pm 1 | t' \leq t, s' < t'\}).$$

This state concentration is well approximated using the Markovian approximation and so the definition af α

$$p_{t+1,s+1} \simeq \int_{q_{ts} \neq \pm 1} W(q_{t+1,s+1=1|q_{ts}}) P(q_{ts})\, dq_{ts} = \exp[n\alpha_{t+1,s+1}(1)] \tag{A.6}$$

In the second passage we used that the result is not sensitive to exclusion of $q' = \pm 1$ from the integral for large n. This is because $\max_{q'}$ is insensitive to its argument at $q' = \pm 1$ unless the initial distribution is sharply peaked around $q = \pm 1$, which is not the case here.

From Markovian processes, the probability that the dynamics starting from $\bar{\sigma}(0)$ comes back for the first time to $\bar{\sigma}(0)$ after l steps without visiting any sign flip of previously visited states is described for large n by

$$\tilde{P}(l) \equiv Prob(\{q_{1,0} \neq \pm 1\}, \{q_{2,s} \neq \pm 1 | s = 0, 1\}, \ldots, \{q_{l-1,s} \neq \pm 1 | s = 0, 1, \ldots, l-2\}, q_{l,0} = 1) = \tag{A.7}$$

$$= (1-2p_{1,0})\prod_{s=0}^{1}(1-2p_{2,s})\cdots\prod_{s=0}^{l-2}(1-2p_{l-1,s})p_{l,0} =$$

$$= p_{l,0}\exp\left(\sum_{t=1}^{l-1}\sum_{s=0}^{t-1}\ln(1-2p_{t,s})\right)$$

note that the factor $\prod_{s=0}^{t-1}(1-2p_{t,s})$ describes the probability that the state makes a transition at time t to a state distinct from $\{\pm\bar{\sigma}(s)|s=0,1,\ldots,t-1\}$. $p_{l,0}$ describes the probability of coming back to the initial state $\bar{\sigma}(0)$ after l steps.
Then, the probability that a certain state, say $\bar{\sigma}(0)$, belongs to a cycle of length l is given for large n by

$$P(l) = \begin{cases} \tilde{P}(l) & (odd\ l) \\ \\ \tilde{P}(l) + \tilde{P}(l/2) & (even\ l) \end{cases}$$

In the case of odd l the result is found, in the other case of even l there are different

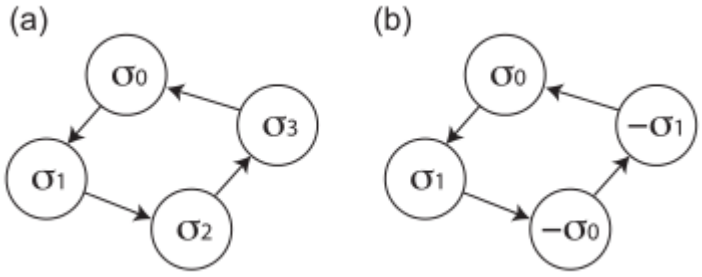


Figure A.2: There are two kinds of limit cycles if the cycle length l is even. (a) In the first kind of cycles, the cycle closes without ever visiting the sign flip of previously visited states. (b) In the second kind of cycles, the state first makes a transition to the sign flip of the initial state after $l/2$ steps, i.e., $\sigma(l/2) = -\sigma(0)$. If this happens, the cycle must close after l steps. [6]

contributions. The first contribution is from cycles that close without ever visiting the sign flip of their history, the second contribution is from from cycles that involves a transition at $l/2$ to the sign flip of their initial state, and this guarantees that the cycles closes in l steps.

The next (and the final) step is to evaluate the state concentration probability $p_{t,s}$. The initial state concentration probabilities are simply given by

$$p_{l,0} \simeq 2^{-n} \equiv p_{init}$$

in the large n and l limit. On the other hand, the quasistationary value for $p_{t+l,t}$ is

$$p_\infty \equiv \lim_{t\to\infty} p_{t+l,t} = \exp[n\alpha(1)]$$

for any $l \geq 1$, where $\alpha(1) = -0.46$. The state concentration quickly converges to the initial value $p_{init} \simeq \exp(-0.69n)$ to the asymptotic value $p_\infty \simeq \exp(-0.46n)$. We can then do the following approximation

$$\tilde{P}(l) = p_{init}\exp\left[\sum_{t=1}^{l-1}\sum_{s=0}^{t-1}\ln(1-2p_\infty) + O\left(2t_c l\frac{p_\infty - p_{init}}{1-2p_\infty}\right)\right] \simeq$$

$$\simeq p_{init} \exp\left[\frac{l^2}{2}\ln(1-2p_\infty)\right] = p_{init}\exp\left(-\frac{l^2}{\tau^2}\right)$$

where $t_c \simeq 5$ and $\tau \equiv \sqrt{-2/\ln(1-2p_\infty)}$ is the characteristic cycle length that grows exponentially with the system size. The approximation used assumes

$$4t_c \frac{p_\infty - p_{init}}{-(1-2p_\infty)\ln(1-2p_\infty)} \ll l$$

$$and \quad l \ll \frac{1-2p_\infty}{2t_c(p_\infty - p_{init})}$$

defines the range of l, that is roughly $10 \ll l \ll \exp(0.46n)/(2t_c)$ at $n > 10$. Then, is a well definition $\tau \simeq \exp(0.23n)$ in this range.

The probability of observing a cycle of length l is given by $P(l)/(lZ)$ with a normalization factor $Z = \sum_{l=1}^{2^n} P(l)/l$, where the probability is divided by l to provide the cycle length probability since all states within a cycle share the same cycle length, so Z is the probability of a state belonging to a cycle. The cumulative distribution of cycle length is similarly obtained by

$$F(l) \equiv \frac{1}{Z}\sum_{l'=1}^{l}\frac{P(l')}{l'} \simeq \frac{1}{Z}\left[\int_1^l \frac{\tilde{P}(l')}{l'}\,dl' + \int_1^{l/2}\frac{\tilde{P}(l')}{2l'}\,dl'\right]$$

The comparison of $F(l)$ with the numerical results is shown in Fig.A.5(b), the discrepancy

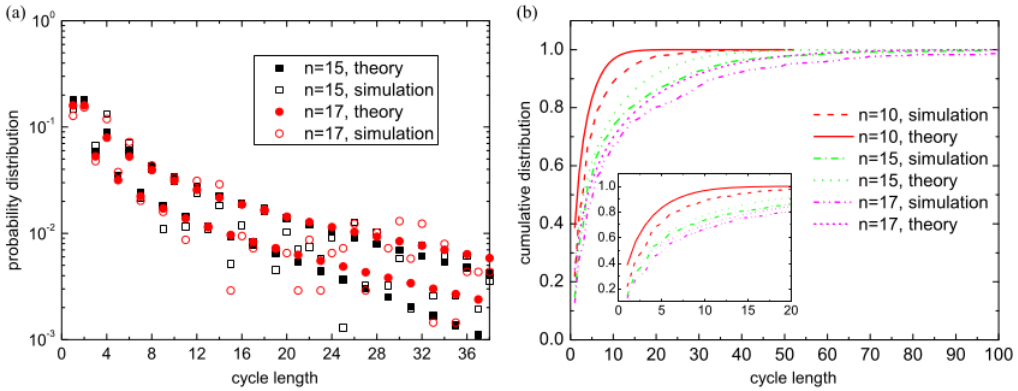


Figure A.3: (Color online) (a) Probability distribution of cycle lengths. (b) Cumulative distribution of cycle lengths. The numerical data is obtained from 1000 samples for $n = 10$, 500 samples for $n = 15$, and 200 samples for $n = 17$. The inset shows an enlarged view at small cycle length. [6]

tends to become small for larger n.
The first moment and the second moment of the distribution can be computed analytically as

$$< l >= \frac{4\sqrt{\pi}\tau[1-erf(1/\tau)]}{3\int_{1/\tau^2}^{\infty}\frac{e^{-t}}{t}\,dt}$$

$$< l^2 >= \frac{2\tau^2 e^{-1/\tau^2}}{\int_{1/\tau^2}^{\infty}\frac{e^{-t}}{t}\,dt}$$

where $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$ and $\int_{1/\tau^2}^{\infty}\frac{e^{-t}}{t}\,dt \simeq -\gamma_E - \alpha(1)n$ in the large n limit, where $\gamma_E = 0.5772$ is the Euler constant.

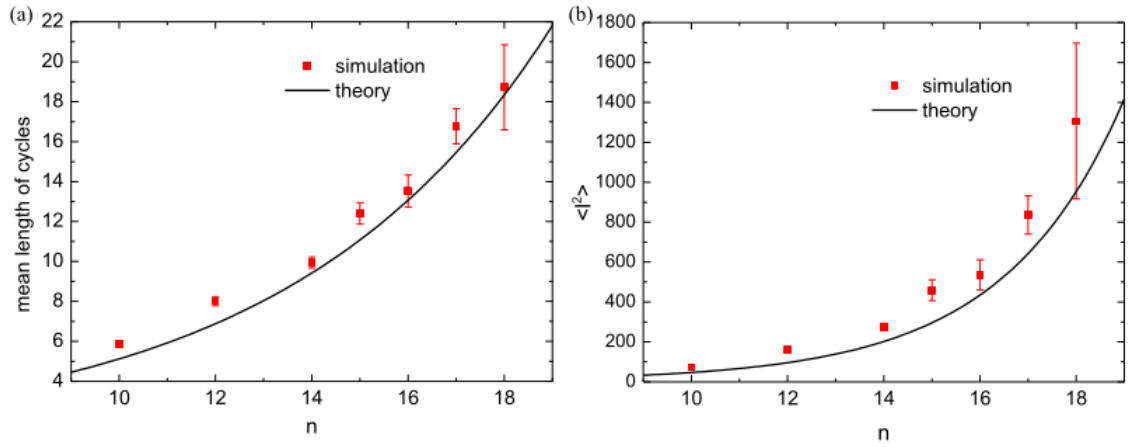


Figure A.4: (Color online) The first (mean) and second moment of the cycle length distribution. Theoretical predictions and numerical simulations are compared. The results are averaged over many random realizations of the networks (from 1000 samples for $n = 10$ to 100 samples for $n = 18$). [6]

At last, another interesting quantity is the number of the attractive states N_{att} belonging to all cycles (i.e. a cycle of length l has l attractive states), which is expected to grow exponentially with the network size n. Theoretically this quantity is

$$N_{att} = 2^n \sum_{l=1}^{2^n} P(l)$$

and can be quantified as the grow rate (entropy density) $s = \lim_{n\to\infty} \frac{1}{n} \ln N_{att}$ In the large n limit, we obtain $s = -\alpha(1)/2$ which is compared with numerical results and as n increases, s decreases approaching the asymptotic limit 0.2277.

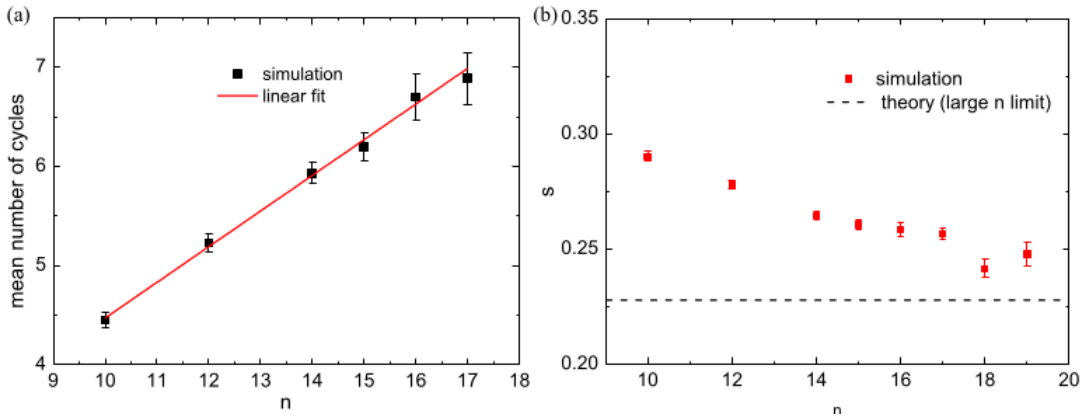


Figure A.5: (Color online) (a) Linear dependence of the number of cycles on network size n. (b) Entropy density of the attractive states defined by $s = \frac{1}{n} \ln N_{att}$. As n increases, the numerical data approach the theoretical prediction. [6]

Appendix B

Appendix 2: Maximum Entropy Models

Systems with many degrees of freedom have a dauntingly large number of states, which grows exponentially with the system's size, a phenomenon called *curse of dimensionality*. For this reason, getting a good estimate of $P(\bar{\sigma})$ from data can be impractical. The *principle of maximum entropy* is a technique for dealing with this problem by assuming a model that is as random as possible, but that agrees with some average observable of the data.
The real distribution $P_r(\bar{\sigma})$ is approximated by a model distribution $P_m(\bar{\sigma})$ that maximizes the Gibbs formula for the entropy:

$$S[P_m(\bar{\sigma})] = -\sum_{\bar{\sigma}} P_m(\bar{\sigma}) \ln P_m(\bar{\sigma}) \tag{B.1}$$

that satisfies

$$\langle \mathcal{O}_a(\bar{\sigma}) \rangle_m = \langle \mathcal{O}_a(\bar{\sigma}) \rangle_r \tag{B.2}$$

where $\mathcal{O}_a$ are a set of observables of the system and the brackets are the average taken with the correspondent distribution (model or real).
The key point is that often average observables $\langle \mathcal{O}_a(\bar{\sigma}) \rangle_r$ can be estimated accurately from the data, even when the whole distribution $P_r(\bar{\sigma})$ cannot. Using the technique of Lagrange multipliers, one can write the explicit form of the model distribution:

$$P_m(\bar{\sigma}) = \frac{1}{Z} e^{\sum_a \beta_a \mathcal{O}_a(\bar{\sigma})}$$

where β_a are the Lagrange multipliers associated to the constraints and constitute the fitting parameters of the model. For example, when the maximum entropy model (MEM) in constrained only by the mean value of the energy, $\mathcal{O}(\bar{\sigma}) = -E(\bar{\sigma})$, we recover the Boltzmann distribution, $P_m(\bar{\sigma}) = Z^{-1} e^{-\beta E(\bar{\sigma})}$, where $\beta = (k_B T)^{-1}$.
More generally, the exponential form of the distribution to find suggests to define the energy as

$$E(\bar{\sigma}) = -\sum_a \beta_a \mathcal{O}_a(\bar{\sigma})$$

There exists a unique set of Lagrange multipliers that satisfies all the constraints, but finding them is a computationally difficult inverse problem. Solving the inverse problem is equivalent to minimizing the Kulballack-Leibler divergence between the real and the model distribution, defined as

$$D_{KL}(P_r \| P_m) = \sum_{\bar{\sigma}} P_r(\bar{\sigma}) \ln \frac{P_r(\bar{\sigma})}{P_m(\bar{\sigma})} \tag{B.3}$$

or qeuivalently, to maximizing the log-likelihood $\mathcal{L}$ that the experimental data was produced by the model:

$$\mathcal{L} = \ln \prod_{a=1}^{M} P_m(\bar{\sigma}^a) = M \sum_{\bar{\sigma}} P_r(\bar{\sigma}) \ln P_m(\bar{\sigma}) = M\{S[P_r] - D_{KL}(P_r \| P_m)\}$$

where, by definition, $P_r(\bar{\sigma}) = (1/M) \sum_{a=1}^{M} \delta_{\bar{\sigma},\bar{\sigma}^a}$. In fact one has:

$$\frac{\partial D_{KL}(P_r \| P_m)}{\partial \beta_a} = \langle \mathcal{O}_a \rangle_m - \langle \mathcal{O}_a \rangle_r$$

This explicit expression of the derivatives suggests to use a gradient descent algorithm, with the following update rules for the model parameters:

$$\beta_a \leftarrow \beta_a + \eta(\langle \mathcal{O}_a \rangle_r - \langle \mathcal{O}_a \rangle_m)$$

where η is a small constant, the *learning rate*. The inverse problem is in fact broken down into two tasks: estimating the mean observables within the model distribution for a given set of parameters β_a and then implementing an update rule that will converge to the right parameter.

Appendix C

Appendix 3: Linear Relation between Energy and Entropy (using Zipf's Law)

Let us denote by $\bar{\sigma}$ the state of the system. Generally, $\bar{\sigma}$ is a multi-dimensional variable $\bar{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ in a M-dimensional space ($M \geq N$) and σ_i can be a spin, a letter in a word, the spiking activity of a neuron, an amino acid in a peptide chain, or the vector velocity of bird in a flock.
Denoted $P(\bar{\sigma})$ the probability of find the system in the state $\bar{\sigma}$, formally we can define this probability as a Boltzmann distribution:

$$P(\bar{\sigma}) = \frac{1}{Z} e^{-E(\bar{\sigma})/k_B T} \tag{C.1}$$

where k_B is Boltzmann's constant and Z the partition function. Without loss of generality we can set the temperature $k_B T = 1$, and Z to 1, which leads to the following definition for the energy:

$$E(\bar{\sigma}) = -\log P(\bar{\sigma}) \tag{C.2}$$

With the availability of large datasets, it now seems possible to construct $P(\bar{\sigma})$ directly from the data, and to take the corresponding energy function $E(\bar{\sigma})$ seriously as a statistical mechanics problem. In this section we explore the consequences of that idea, by showing the equivalence between Zipf's law of language and the critical properties of the associated statistical mechanics model.
There is a very old observation of a power law in a biological system, and this is Zipf's law, first observed by Auerbach in 1913. Zipf's law refers to the distribution over states of the system, in the same way that the Boltzmann distribution describes the distribution over states of an equilibrium system. We can think of the state of the system as being a single word $\bar{\sigma}$, and as texts or conversations proceed they sample many such states. If one orders (ranks) words $\bar{\sigma}$ by their decreasing frequency $P(\bar{\sigma})$, Zipf's law states that the frequency of words $P(\bar{\sigma})$ decays as the inverse of their rank $r(\bar{\sigma})$:

$$P(\bar{\sigma}) \propto \frac{1}{r(\bar{\sigma})} \tag{C.3}$$

This can be corrected either by introducing a cutoff corresponding to a finite vocabulary, or by slightly modifying the law to $P = \frac{r^{-\alpha}}{\zeta(\alpha)}$, with $\alpha > 1$ and $\zeta(\alpha)$ is Riemann's zeta

function. Since its introduction in the context of language, Zipf's law has been observed in all branches of science, but has also attracted a lot of criticism, essentially for the same reasons as other power laws, but also because of the controversial claim by Zipf himself that his law was characteristic of human language.
Despite all our concerns, Zipf's law is, in a certain precise sense, a signature of criticality.
Consider the density of states, obtained just by counting the number of states in a small window δE. The density of states is defined as follows:

$$\rho_{\delta E}(E) = \frac{1}{\delta E} \sum_{\bar{\sigma}} \mathbb{I}[E < E(\bar{\sigma}) < E + \delta E] \tag{C.4}$$

where $\mathbb{I}$ is the indicator function. In the thermodynamic limit the density of states is the exponential of the entropy, so

$$S(E) \equiv \log \rho_{\delta E}(E) = Ns + s_1 \tag{C.5}$$

where s_1 is sub-extensive, that is $\lim_{N\to\infty} s_1/N = 0$. For real data and finite N , the choice of the bin size δE can be problematic, and it is useful to consider instead the cumulative density of states:

$$\mathcal{N}(E) = \sum_{\bar{\sigma}} \mathbb{I}[E(\bar{\sigma}) < E] = \int_{-\infty}^{E} \rho_{\delta E=0}(E)\, dE' \tag{C.6}$$

In the case of large systems, the integral is dominated by the maximum of the integrand:

$$\mathcal{N}(E) = N \int_{-\infty}^{E/N} d\varepsilon' \exp[N(s(\varepsilon') + s_1/N] \sim e^{Ns(\varepsilon)} \tag{C.7}$$

so we got

$$\log \mathcal{N}(E) \sim Ns(E/N) = S(E) \tag{C.8}$$

The rank $r(\bar{\sigma})$ previously defined is the cumulative density of states at the energy $E(\bar{\sigma})$:

$$r(\bar{\sigma}) = \mathcal{N}[E = E(\bar{\sigma})] \tag{C.9}$$

For large systems then in general we expect that

$$S(E(\bar{\sigma})) \sim \log r(\bar{\sigma}) \tag{C.10}$$

Using Zipf's law

$$-\log P(\bar{\sigma}) = \alpha \log r(\bar{\sigma}) + \log \xi(\alpha) \tag{C.11}$$

$$S(E) = \frac{E}{\alpha} + \dots \tag{C.12}$$

In words, Zipf's law for a very large system is equivalent to the statement that the entropy is exactly a linear function of the energy.

We will use this result further analyzing the activity of a neuronal network, searching some clues of criticality behind the pattern of time activity. A perfect linear relation between energy and entropy is very unusual, to see why let's recall the canonical partition function:

$$Z(T) = \sum_{\bar{\sigma}} e^{-E(\bar{\sigma})/k_B T} \tag{C.13}$$

Where is introduced a fictitious temperature T (that can be posed as $k_B T) = 1$. We have

$$Z(T) = \int \rho(E) e^{-E/k_B T}\, dE \sim \int exp[N(s(\varepsilon) - \varepsilon/k_B T)]\, d\varepsilon \tag{C.14}$$

For large N, the integral is dominated by the largest term of the integrand $f(E) = N(s(\varepsilon)) - \varepsilon/k_BT$ which is the following point:

$$0 = \frac{df(E)}{dE} = N(ds(E) - d\varepsilon/k_BT) \tag{C.15}$$

from which we obtain

$$\frac{ds}{d\varepsilon} = \frac{1}{k_BT} \tag{C.16}$$

In the special case of Zipf's law, it is found that $ds/d\varepsilon = 1/\alpha \; \forall E$. This means that $k_BT = \alpha$ is a critical point, for any $k_BT < \alpha$ the system freezes into a ground state of zero energy and zero entropy, while for $k_BT > \alpha$ the systems explores higher energies with ever higher probabilities.
It is found that for Zipf's law $d^2S/dE^2 = 0 \; \forall E$, making this law a strong signature of criticality. A consequence of this is that the entropy is sub-extensive below the critical point, $S/N \to 0$.

Appendix D

Examples of power-law distributions

The ubiquity of power-law behaviour in the natural world has led many scientists to wonder whether there is a single, simple, underlying mechanism linking all these different systems together. Several candidates for such mechanisms have been proposed, going by-names like "self-organized criticality" and "highly opti- mized tolerance". However, the conventional wisdom is that there are actually many different mechanisms for producing power laws and that different ones are applicable to different cases. Now we see various examples of power-laws in real systems [20]:

Word frequency:

Estoup [21] observed that the frequency with which words are used appears to follow a power law, and this observation was famously examined in depth and confirmed by Zipf [22]. Panel (a) of Fig.D.1 shows the cumulative distribution of the number of times that words occur in a typical piece of English text, in this case the text of the novel Moby Dick by Herman Melville. Similar distributions are seen for words in other languages

Citations of scientific papers:

As first observed by Price [23], the numbers of citations received by scientific papers appear to have a power-law distribution. The data in panel (b) are taken from the Science Citation Index, as collated by Redner [24], and are for papers published in 1981. The plot shows the cumulative distribution of the number of citations received by a paper between publication and June 1997.

Web hits:

The cumulative distribution of the number of "hits" received by web sites (i.e., servers, not pages) during a single day from a subset of the users of the AOL Internet service.

Copies of books sold:

The cumulative distribution of the total number of copies sold in America of the 633 bestselling books that sold 2 million or more copies between 1895 and 1965.

Telephone calls:

The cumulative distribution of the number of calls received on a single day by 51 million users of AT&T long distance telephone service in the United States. After Aiello et al. [25]. The largest number of calls received by a customer in that day was 375 746, or about 260 calls a minute (obviously to a telephone number that has many people manning the phones). Similar distributions are seen for the number of calls placed by users and also for the numbers of email messages that people send and receive [26, 27].

Magnitude of earthquakes:

The cumulative distribution of the Richter (local) magnitude of earthquakes occurring in California between January 1910 and May 1992, as recorded in the Berkeley Earthquake Catalog. The Richter magnitude is defined as the logarithm, base 10, of the maximum amplitude of motion detected in the earthquake, and hence the horizontal scale in the plot, which is drawn as linear, is in effect a logarithmic scale of amplitude. The power law relationship in the earthquake distribution is thus a relationship between amplitude and frequency of occurrence.

Diameter of moon craters:

The cumulative distribution of the diameter of moon craters. Rather than measuring the (integer) number of craters of a given size on the whole surface of the moon, the vertical axis is normalized to measure number of craters per square kilometer, which is why the axis goes below 1, unlike the rest of the plots, since it is entirely possible for there to be less than one crater of a given size per square kilometer. After Neukum and Ivanov [28].

Intensity of solar flares:

The cumulative distribution of the peak gamma-ray intensity of solar flares. The observations were made between 1980 and 1989 by the instrument known as the Hard X-Ray Burst Spectrometer aboard the Solar Maximum Mission satellite launched in 1980. The spectrometer used a CsI scintillation detector to measure gamma-rays from solar flares and the horizontal axis in the figure is calibrated in terms of scintillation counts per second from this detector.

Intensity of wars:

The cumulative distribution of the intensity of 119 wars from 1816 to 1980. Intensity is defined by taking the number of battle deaths among all participant countries in a war, dividing by the total combined populations of the countries and multiplying by 10 000. For instance, the intensities of the First and Second World Wars were 141.5 and 106.3 battle deaths per 10 000 respectively. The worst war of the period covered was the small but horrifically destructive Paraguay-Bolivia war of 1932–1935 with an intensity of 382.4.

Wealth of the richest people:

The cumulative distribution of the total wealth of the richest people in the United States.

Frequencies of family names:

Cumulative distribution of the frequency of occurrence in the US of the 89 000 most common family names, as recorded by the US Census Bureau in 1990. Similar distributions

are observed for names in some other cultures as well but not in all cases. Korean family names for instance appear to have an exponential distribution [29].

Populations of cities:

Cumulative distribution of the size of the human populations of US cities as recorded by the US Census Bureau in 2000.

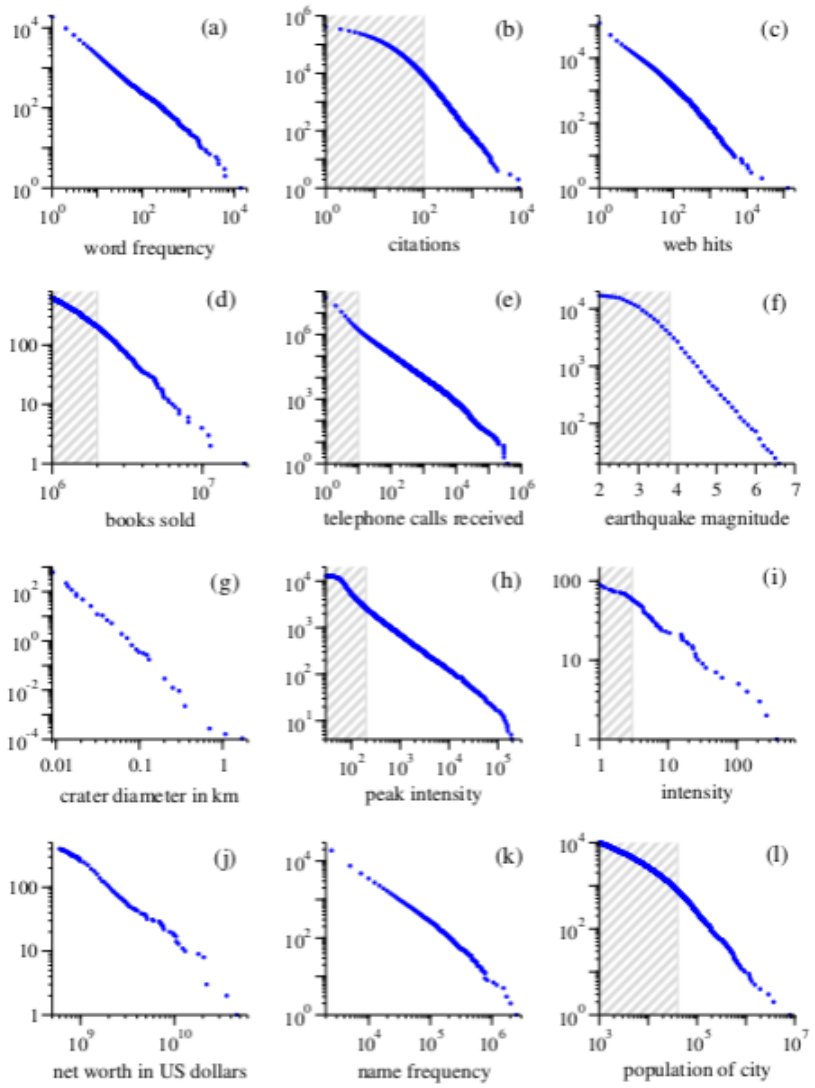


Figure D.1: Cumulative distributions or "rank/frequency plots" of twelve quantities reputed to follow power laws. (a) Numbers of occurrences of words in the novel Moby Dick by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of hits on web sites by 60 000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a (g) Diameter of craters on the moon. Vertical axis is measured per square kilometer. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.power law even though the horizontal axis is linear.

Bibliography

[1] R. G. Ingalls, M. D. Rossetti, J. S. Smith and B. A. Peters, eds. "More "normal" than normal: scaling distributions and complex systems"

[2] Thierry Mora, William Bialek "Are biological system posed at criticality?", Springer Science 2011

[3] Alireza Goudarzi, Christof Teuscher, Natali Gulbahce, Thimo Rohlf "Emergent criticality through adaptive information processing in boolean networks", Portland University PDXScholar 3-1-2012

[4] Johnatan Touboul, Alain Destexhe "Power-law statistics and universal scaling in the absence of criticality"

[5] B. Deridda, Y. Pomeau "Random networks of automata: a simple annealed approximation, Europhysics letters, 1 (2) pp. 45-49 (1986)

[6] Taro Toyoizumi, Haiping Huang "Structure of attractors in randomly connected networks", Physical Review E 91, 032802 (2015)

[7] N. Fusco, P. Marcellini, C. Sbordone "Analisi Matematica due", Liguori Editore

[8] Enzo Orlandini "Meccanica Statistica II", handouts for the course of statistical mechanics of the Master of Science in Physics for the Universitá of Padova

[9] R. Griffiths, Phase transitions and critical phenomena, vol. 1, (ed. Domb e Green, Academic Press, New York 1972)

[10] Luca Peliti, Appunti di meccanica statistica, (Bollati Boringhieri 2003)

[11] Robert E. Marc, "The Structure of Vertebrate Retinas"

[12] Flock of birds images: http://alaska.usgs.gov/science/kasatochi/

[13] Ole Peters & J. David Neelin, "Critical phenomena in atmospheric precipitation", Nature Physics, Revised, 04/2006

[14] John M. Beggs and Dietmar Plenz, "Neuronal Avalanches in Neocortical Circuits", The Journal of Neuroscience, December 3, 2003 • 23(35):11167–11177 • 11167

[15] David A. Kessler, Herbert Levine, "Generic Criticality in Ecological and Neuronal Networks", arXiv 1508.02414v1 10 August 2015

[16] Giuseppe De Marco, "Analisi Matematica Due", Zanichelli

[17] Aaron Clauset, Cosma Rohilla Shalizi, M. E. J. Newman, "Power-law distributions in empirical data", arXiv:0706.1062 2 February 2009

[18] Muniruzzaman, A. N. M., 1957, Bulletit on of the Calcutta Statistical Association 7, 115.

[19] Hill, B. M., 1975, Annals of Statistics 3, 1163.

[20] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law", arXiv 0412004v3 29 march 2006

[21] J. B. Estoup, Gammes Stenographiques. InstitutStenographique de France, Paris (1916)

[22] G. K. Zipf, Human Behaviour and the Principle of Least Effort. Addison-Wesley, Reading, MA (1949).

[23] D. J. de S. Price, Networks of scientific papers. Science 149, 510–515 (1965).

[24] S. Redner, "How popular is your paper? An empirical study of the citation distribution." Eur. Phys. J. B 4, 131–134 (1998).

[25] W. Aiello, F. Chung, and L. Lu, A random graph model for massive graphs. In Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, pp. 171–180, Association of Computing Machinery, New York (2000).

[26] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Scale-free topology of e-mail networks. Phys. Rev. E 66, 035103 (2002).

[27] B. A. Huberman and L. A. Adamic, Information dynamics in the networked world. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai (eds.), Complex Networks, num- ber 650 in Lecture Notes in Physics, pp. 371–398, Springer, Berlin (2004).

[28] G. Neukum and B. A. Ivanov, Crater size distributions and impact probabilities on Earth from lunar, terrestial-planet, and asteroid cratering data. In T. Gehrels (ed.),Hazards Due to Comets and Asteroids, pp. 359–416, University of Arizona Press, Tucson, AZ (1994).

[29] B. J. Kim and S. M. Park, Distribution of Korean family names. Preprint cond-mat/0407311 (2004).

[30] H. Cramér and M. Leadbetter, "Stationary and Related Stochastic Processes," John Wiley, New York, 1966.

[31] S. Ross, "Stochastic Processes," Second Edition, John Wiley, New York, 1996.

[32] S. Ross, “Probability Models for Computer Science”, Academic Press, 2002.

[33] NAVAL POSTGRADUATE SCHOOL of Monterey, California, ”Simulation of nonhomogeneous Poisson processes by thinning”,P. A. W. Lewis & G. S. Shedler, June 1978.

[34] http://www.treccani.it/enciclopedia/menti-collettive

[35] G. Zambotti, ”Introduzione alla MECCANICA STATISTICA”, handouts for the course of nuclear and theoretical physics of the Master of Science in Physics for the Universitá of Pavia

[36] M. Plischke and B. Bergersen, Equilibrium Statistical Physics, (Prentice-Hall,1989).

[37] George A. Baker Jr., Quantitative theory of Critical Phenomena, (Academic Press 1990).

[38] R. Griffiths, Phase transitions and critical phenomena, vol. 1, (ed. Domb e Green, Academic Press, New Yor 1972).

[39] M. Kardar and A. N. Berker Commensurate-Incommensurate Phase Diagrams for Overlayers from a Helical Potts Model Phys. Rev. Lett. 48, 1552 (1982).

[40] KG Wilson, Problems in physics with many scales of length. Sci Am 241, 158–179 (1979).

[41] JP Sethna, Statistical Mechanics: Entropy, Order Parameters, and Complexity (Oxford University Press, Oxford, 2006).

[42] G Parisi, Statistical Field Theory (Addison–Wesley, Redwood City CA, 1988).

[43] J Guckenheimer and P Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields (Springer–Verlag, New York, 1983).

[44] P. Bak, C. Tang, K. Wiesenfeld, et al., Physical review letters 59, 381 (1987).

[45] P. Bak, How nature works (Oxford university press Oxford, 1997).

[46] W. L. Shew and D. Plenz, The neuroscientist 19, 88 (2013).

[47] D. R. Chialvo, Nature physics 2, 301 (2006).

[48] Samir Suweis, ”DINAMICA STOCASTICA DI MODELLI PER ECOSISTEMI.”, Master’s Thesis in physics, Academic Year 2007/08

[49] Maritan A. ”Corso di fisica dei sistemi complessi”.

[50] S. Chandrasekhar, ”Stochastic problems in physics and astronomy” Rev. Mod. Phys., 15(1):1–89, Jan 1943.

[51] Dante R. Chialvo, "Emergent complex neural dynamics", Nature Physics 1 October 2010

[52] Jordi Hidalgo, Jacopo Grilli, Samir Suweis, Miguel A. Muñoz, Jayanth R. Banavar, and Amos Maritan "Information-based fitness and the emergence of criticality in living systems", PNAS 111(28), 10095–10100 (2014)

[53] Gasper Tkacik, Thierry Mora, Olivier Marre, Dario Amodei,Michael J. Berry II, and William Bialek, "Thermodynamics for a network of neurons: Signatures of criticality", arXiv 1407.5946v1 22 July 2014

[54] John M. Beggs and Nicholas Timme "published: "Being critical of criticality in the brain", Frontiers in Physiology 07 June 2012 doi: 10.3389/fphys.2012.00163

[55] Stephen Jay Gould & Niles Eldredge, "Punctuated equilibrium comes of age", NATURE - VOL 366 - 18 NOVEMBER 1993

[56] S Schnabel, DT Seaton, DP Landau, and M Bachmann, Microcanonical entropy inflection points: Key to systematic understanding of transitions in finite systems. Phys Rev E 84, 011127 (2011).

[57] P.W. Anderson "More is different", Science, New Series, Vol. 177, N. 4047 (Aug. 4 1972), 393-396

[58] Tamás Vicsek, Anna Zafeiris "Collective motion" , PhysicsReports517(2012)71–140

[59] MEJ Newman (2005) Power laws, Pareto distributions and Zipf's law, Contemporary Physics, 46:5, 323-351, DOI: 10.1080/00107510500052444

[60] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi Raffaele Santagati, Fabio Stefanini and Massimiliano Viale "Scale-free correlations in starling flocks" PNAS June 29, 2010 vol. 107 no. 26 11865 11870

[61] Alessandro Attanasi, Andrea Cavagna, Lorenzo Del Castello, Irene Giardina, Stefania Melillo, Leonardo Parisi, Oliver Pohl, Bruno Rossaro, Edward Shen, Edmondo Silvestri, and Massimiliano Viale, "Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms". Phys. Rev. Lett. 113, 238102 – Published 1 December 2014

[62] J. J. Binney, N. J. Dowrick, A. J. Fisher, M. Newman, "The Theory of Critical Phenomena: An Introduction to the Renormalization Group", Oxford University Press, Inc. New York, NY, USA 1992 ISBN:0198513933 9780198513933

[63] Dante R. Chialvo, "Emergent complex neural dynamics", Nature Physics 6, 744–750 (2010)

[64] Per Bak, "How Nature Works: The Science of Self-Organised Criticality", New York, NY: Copernicus Press 1996 Cloth: ISBN 0-387-94791-4