

Article

Entropic Divergence and Entropy Related to Nonlinear Master Equations

Tamás Sándor Biró ^{1,*}, †, ‡ , Zoltán Nédá ²  and András Telcs ^{1,3,4,5} 

¹ Wigner Research Centre for Physics, 1121 Budapest, Hungary; Telcs.Andras@wigner.mta.hu

² Department of Physics, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania; zneda@phys.ubbcluj.ro

³ Department of Computer Science and Information Theory, Budapest University of Technology and Economics, 1111 Budapest, Hungary

⁴ Department of Quantitative Methods, University of Pannonia, 8200 Veszprém, Hungary

⁵ VIAS Virtual Institute for Advanced Studies, 1039 Budapest, Hungary

* Correspondence: Biro.Tamas@wigner.mta.hu; Tel.: +36-20-435-1283

† Based on a talk given at JETC, Barcelona, 21–24 May 2019.

‡ External faculty, Complex Science Hub, 1080 Vienna, Austria.

Received: 11 September 2019; Accepted: 8 October 2019; Published: 11 October 2019



Abstract: We reverse engineer entropy formulas from entropic divergence, optimized to given classes of probability distribution function (PDF) evolution dynamical equation. For linear dynamics of the distribution function, the traditional Kullback–Leibler formula follows from using the logarithm function in the Csiszár’s f -divergence construction, while for nonlinear master equations more general formulas emerge. As applications, we review a local growth and global reset (LGGR) model for citation distributions, income distribution models and hadron number fluctuations in high energy collisions.

Keywords: entropy; entropic divergence; master equation; reset; preferential growth

1. Motivation

The entropic convergence is being intensively studied alongside functional inequalities.

The log–Sobolev inequality and a family of other related functional inequalities have been developed to control the speed of convergence of stochastic processes on the Euclidian space and on manifolds. The control is given in a form of an upper bound on the speed of convergence in one or other metrics, e.g., Wasserstein distance, L_p -norm or other. The most important development by this approach is the incorporation of the curvature condition, which replaces the condition on the lower bound of the spectral gap. The latter one can be typically used on compact sets while the former allows us to control the convergence on non-compact sets, on spaces of infinite diameter or measure.

The theory of measure concentration (“entropic distance shrinking”) received a new impact from the work of Marton [1] and Talagrand [2]. This initiative later grow into a unified treatment by Otto and Villanyi [3] incorporating mass transport theory and gradient flow on manifolds of measures. All those efforts are focused on finding conditions for the convergence of the process under scrutiny. In these works several extensions of the log-Sobolev inequality, Wasserstein distance and Fisher–Donsker–Varadhan information is established [4].

The present work is dealing with processes for which such methods does not work to our best knowledge. Convergence is ensured in a different way. A metric, in particular a Kullback-type divergence [5] is created in which—thanks to the intimate relation to the structure of the master equation—it can be shown that its time derivative is negative.

For recent developments dealing with discrete Markov chains see also References [6,7] and references therein.

The classical entropy—probability relation, pioneered by Boltzmann, used and discussed by Gibbs, Planck and Shannon [8–12] (just to mention a few), is under steady generalization attempts, e.g., [13–18]. Most of the suggestions are mathematically motivated, formulated in terms of axioms (Khinchine or other) [19,20], sometimes abandoning the additivity property for factorizing probabilities, therefore treating statistical independence in an unorthodox way [21]. The motivation behind doing so lies in the study of complex systems, where complexity reveals itself in long-range correlations causing unusual behavior in the thermodynamical limit. It is common in such phenomena that interface effects, multiplied with a characteristic correlation length, are not smaller than bulk (volume) effects. This ruins the familiar hierarchy valid for big systems in the classical thermodynamics and is mathematically reflected in the non-extensivity property. Hence formulas generalizing the logarithmic function at the core of the entropy formula are necessarily non-extensive. Besides this, some approaches keep the logarithm but use a function of the probability in the formula. For example, in the field of sensor data analysis, the Deng entropy is in use [22–24].

By the generalized entropy formula, important properties, like non-negativity, expansibility, convexity are still to be satisfied [25]. It is, however, hard to convince about one or the other “next to simplest” generalization without investigating physical phenomena which realize or—if we are lucky—even suggest preferring a given formula against others [26,27]. First, in a deductive way someone suggests a formula and then checks its fundamental properties before applying to a number of physical observations. Any suggested new entropy formula must include the conventional one as a limiting case, since this has already the largest observational support. A second way to select a favorite formula may rely on starting with the study of a given class of complex system behavior and adjust the classical definition accordingly. Most importantly, one expects that the formula commands the entropy to increase in a closed system, under the given microscopic dynamics already known or assumed.

Although both ways are legitimate, in this paper we follow the second one. Given an evolution dynamics for the microstate probabilities, we first seek for an expedient formula for the entropic divergence [28]: a non-negative measure between two probability distributions (in a continuous model probability distribution functions (PDFs) which shrinks during the dynamical evolution. Possibly for two arbitrary initial distributions. Or at least between an arbitrary initial distribution and the stationary distribution. Once this convergence criterion is fulfilled, we study the entropic divergence from the uniform distribution. In special cases, we shall realize that the entropic divergence from the uniform distribution can be written as proportional to a difference of the same functionality on the uniform and on the investigated distribution. Finally this functional will be proposed as a candidate for an entropy formula. For linear master equations describing the evolution of probability, this turns out to be the Boltzmann entropy formula, while for other master equations, different formulas result. In particular, for a fractal power dependence on the probability, the Tsallis entropy formula emerges. Our method described above does not assume detailed balance while delivering the proof, in this way it is valid in a much broader class of dynamics than Boltzmann’s famous H-theorem.

It is often assumed that a non-exponential energy distribution would be the sign for non-extensive entropy. This assumption is, however, mistaken. Both a non-extensive entropy formula with a linear constraint on the energy and a traditional, logarithmic entropy formula with a constraint on another function of the energy lead to non-exponential PDFs in equilibrium. Moreover, several physical systems tend to a statistical behavior with a non-exponential stationary distribution without reaching equilibrium (open systems). In the framework of local growth and global reset model (LGGR) we demonstrate that far from the detailed balance state stationary distributions show a thermodynamical behavior, akin to the fluctuation–dissipation theorem [29]. Examples in income distribution, popularity measured by citations and hadron multiplicities in high-energy collisions demonstrate the flexibility and usefulness of the LGGR model.

As part of this presentation, we repeat some formulas published already earlier [26,28]. Our purpose is to provide the reader with a self-contained chain of thoughts, followable without

further reading. For helping to concentrate on the distinction, however, we summarize here briefly the main development relative to our earlier publications. The general proof of deriving of the shrinking of the entropic divergence is now given for two arbitrary distributions, contrary to the former and more traditional derivation when one of the distributions is the stationary one. We also present some details on why, in the generalized Markovian dynamics, a similar proof based on the convexity property of the definition of the entropic divergence suffices only for an approach to the stationary distribution. Arguments for the generalization to nonlinear probability dynamics are shortly mentioned, but the applications in the framework of our LGGR model is again linear, concentrating on the processes with a stationary state without detailed balance.

2. Shrinking Entropic Divergence

Entropic divergence (sometimes cited as “entropic distance”, however, it is usually not a symmetric measure, nor does it satisfy any triangle inequality) is a non-negative real-valued functional of two probability distributions, $\rho[P, Q] \geq 0$. We also want that whenever this measure is zero, then the equality of the two distributions follows: $\rho[P, Q] = 0 \Rightarrow \forall n P_n = Q_n$. We add a further requirement, related to the evolution dynamics of the probabilities. This irreversibility property expresses that starting with an arbitrary distribution at time $t = 0$, it will approach the stationary distribution, so their entropic divergence shrinks. To begin with we consider Csiszar’s f -divergence formula [30,31]

$$\rho[P, Q] = \sum_n Q_n f(P_n/Q_n), \quad (1)$$

as a class of entropy divergence formulas investigated here. Applying the Jensen inequality [32,33] to this formula one obtains

$$\rho[P, Q] \geq f\left(\sum_n P_n\right) = f(1) \quad (2)$$

for $f'' > 0$. The non-negativity property is hence satisfied by any function in the Csiszar formula with the correct second derivative and $f(1) = 0$.

Next we review a linear dynamics for the probability evolution,

$$\dot{P}_n = \sum_m (w_{nm} P_m - w_{mn} P_n), \quad (3)$$

adjusted to conserve the normalization, i.e., $\sum_n \dot{P}_n = 0$. Here the transition rates from the state m to n , $w_{nm} \geq 0$, define the probability evolution.

Now we are interested in the evolution of the entropic divergence, while both distributions—initially arbitrary—evolve governed by the common transition rates, w_{nm} . This description includes complex systems, where transitions can happen between non-neighboring, in the index space far away states, too. We make no restriction on $n - m$. Now using the definition in Equation (1) the evolution of the entropic divergence shows the following time derivative

$$\dot{\rho} = \sum_n \left(\dot{P}_n f'(\xi_n) + \dot{Q}_n [f(\xi_n) - \xi_n f'(\xi_n)] \right), \quad (4)$$

with the notation $\xi_n = P_n/Q_n$. Substituting the dynamical master Equation (3) we obtain

$$\dot{\rho} = \sum_{nm} \left(w_{nm} Q_m [\xi_m f'(\xi_n) + f(\xi_n) - \xi_n f'(\xi_n)] - w_{mn} Q_n [f(\xi_n)] \right). \quad (5)$$

In the last term we exchange the summation indices to arrive at

$$\dot{\rho} = \sum_{nm} w_{nm} Q_m [(\xi_m - \xi_n) f'(\xi_n) + f(\xi_n) - f(\xi_m)]. \quad (6)$$

Now a definite statement can be made about the sign of this expression by utilizing the Taylor series remainder theorem in the Lagrange form [34] for the yet unspecified function, $f(\xi)$:

$$f(\xi_m) = f(\xi_n) + (\xi_m - \xi_n) f'(\xi_n) + \frac{1}{2} (\xi_m - \xi_n)^2 f''(c_{nm}), \quad (7)$$

with c_{nm} being an intermediate argument between ξ_n and ξ_m .

Using this knowledge we finally obtain

$$\dot{\rho} = -\frac{1}{2} \sum_{nm} w_{nm} Q_m (\xi_m - \xi_n)^2 f''(c_{nm}). \quad (8)$$

For any definite sign of the second derivative of $f(\xi)$, the above formula also has a definite sign, for $f'' > 0$ functions we have $\dot{\rho} \leq 0$. We note that $f'' > 0$ was the very same condition necessary for the non-negativity property of ρ itself, proven via the Jensen inequality, Equation (2). Consequently, we have proven that two arbitrary initial PDFs evolving via the linear master Equation (3) show a shrinking entropic f-divergence between them. This shrinking of the f-divergence from the stationary distribution to an arbitrary one is a basic property, discussed in the literature on Markov processes (for a review see e.g., [35]).

How far can this result be generalized for nonlinear evolution scenarios? We consider a slight modification using a positive function of the probability, $a(P)$, in the master equation

$$\dot{P}_n = \sum_m (w_{nm} a(P_m) - w_{mn} a(P_n)). \quad (9)$$

Here $a(P)$ can be any positive function, including the most known $a(P) = P$ case. In some other complex system kinetics here powers may occur, as in chemical reaction networks [36], morphogenesis [37] or in analogy to some models of finance dynamics generalizing the Fokker-Planck equation to a nonlinear dependence on the phase space density [38]. Other functions may appear when simulating Boltzmann–Uehling–Uhlenbeck type of finite-state blocking or enhancement factors as in quantum transport models [39]. In such problems, one would apply $a(P) = P/(1 \pm P)$.

In this case, the above presented general proof does not apply. The more general trace formula for the entropic divergence,

$$\rho[P, Q] = \sum_n \sigma(P_n, Q_n), \quad (10)$$

has the rate of change

$$\dot{\rho} = \sum_n \left(\dot{P}_n \frac{\partial \sigma}{\partial P_n} + \dot{Q}_n \frac{\partial \sigma}{\partial Q_n} \right). \quad (11)$$

Replacing the dynamical Equation (9) both for \dot{P}_n and \dot{Q}_n and re-arranging the double summation indices as well as using the notation $\xi_n = a(P_n)/a(Q_n)$ we obtain

$$\dot{\rho} = \sum_{n,m} w_{nm} a(Q_m) \left[\xi_m \left(\frac{\partial \sigma}{\partial P_n} - \frac{\partial \sigma}{\partial P_m} \right) + \left(\frac{\partial \sigma}{\partial Q_n} - \frac{\partial \sigma}{\partial Q_m} \right) \right]. \quad (12)$$

For simplifying we introduce the definition

$$f(\xi) \equiv \xi \frac{\partial \sigma}{\partial P} + \frac{\partial \sigma}{\partial Q} \quad (13)$$

for the arbitrary indexed quantities. Using this assertion the change of the divergence is straightforwardly written as

$$\dot{\rho} = \sum_{n,m} w_{nm} a(Q_m) \left[(\xi_m - \xi_n) \frac{\partial \sigma}{\partial P_n} + f(\xi_n) - f(\xi_m) \right]. \quad (14)$$

In order to apply the proof given in the linear master equation case and recited above, we have to achieve

$$\frac{\partial \sigma}{\partial P_n} = f'(\xi_n) \quad \text{and} \quad \frac{\partial \sigma}{\partial Q_n} = f(\xi_n) - \xi_n f'(\xi_n). \quad (15)$$

Whether it is possible or not, can be decided upon investigating the mixed second partial derivative of $\sigma(P, Q)$, i.e., the integrability condition:

$$\frac{\partial^2 \sigma}{\partial Q_n \partial P_n} = \frac{\partial^2 \sigma}{\partial P_n \partial Q_n}. \quad (16)$$

In our particular case this results in

$$-\xi_n f''(\xi_n) \frac{a'(Q_n)}{a(Q_n)} = -\xi_n f''(\xi_n) \frac{a'(P_n)}{a(Q_n)}. \quad (17)$$

This does not happen, unless $a'(P) = a'(Q)$. This restricts us to the linear case or motivates investigations of non-trace like entropy divergence formulas.

It can be proven, however, that the above more general entropy divergence formula Equation (10), shrinks if we take Q_n as the stationary distribution. The latter is defined by the “total balance” condition

$$0 = \sum_m (w_{nm} a(Q_m) - w_{mn} a(Q_n)). \quad (18)$$

We note that the “detailed balance” condition would annulate each term in the above sum, not only the total result. That classical condition is actually a condition on the transition rates, reflecting microscopic time reversibility, while the total balance (consisting of $n = 1, 2, \dots, W$ equations) simply defines the stationary distribution Q_n .

In order to use the argumentation which holds in the linear case, one has only to satisfy

$$\frac{\partial \sigma}{\partial P_n} = f'(\xi_n), \quad (19)$$

with the modified definition $\xi_n = a(P_n)/a(Q_n)$.

The above equation can be solved for any positive $a(P)$ function, particular solutions are set by the natural constraint $\rho[Q, Q] = 0$. This result, stating that all distributions converge to the stationary one due to such $a(P)$ -dynamics, does not mean that the entropic distance between two non-stationary distribution would continuously shrink during their evolution. They will all reach the stationary state and reduce their entropic divergence to zero only in the final stage.

It is traditional to use the function $f(\xi) = -\ln \xi$ for generating the entropic divergence formula. Indeed in this case $f'(\xi) = -1/\xi$ and $f''(\xi) = 1/\xi^2 > 0$. The usual argumentation behind this choice is a further property, making the core function, $f(\xi)$, in the Csiszar formula additive. For composite systems with subsystems (1) and (2) the probabilities namely factorize if statistical independence holds: $P_n(1 \oplus 2) = P_n(1)P_n(2)$ and similarly for Q_n . This makes $\xi(1 \oplus 2) = \xi(1)\xi(2)$ and for $f(1 \oplus 2) = f(1) + f(2)$ with positive second derivative only this choice remains. Using this starting point the entropic divergence formula, Equation (1), and returning to the $a(P) = P$ case specifies to the Kullback–Leibler definition [5,40]

$$\rho^{(KL)}[P, Q] = \sum_n Q_n \ln \frac{Q_n}{P_n}. \quad (20)$$

This formula contains a so called “cross entropy” term and a pure entropy term associated to the PDF Q , but it is not yet unique what entropy formula would follow from this. To make this last step

we utilize our concept of complexity: the entropic divergence of the uniform distribution, $U_n = 1/W$ for $n = 1, 2, \dots, W$ from Q_n , i.e.,

$$\rho^{(KL)}[U, Q] = \sum_{n=1}^W Q_n \ln(WQ_n) = \ln W + \sum_{n=1}^W Q_n \ln Q_n, \quad (21)$$

equals to a difference, $\rho^{(KL)}[U, Q] = S^{(BG)}[U] - S^{(BG)}[Q]$, with the traditional definition of the Boltzmann–Gibbs entropy

$$S^{(BG)}[Q] \equiv - \sum_n Q_n \ln Q_n. \quad (22)$$

In this case the Kullback–Leibler based complexity measure turns out to be a difference of the Boltzmann entropies, as suggested by the P-linear master equation dynamics as a proper measure. Its generalization, however, is not a difference of more general entropy formulas. The entropic divergence formula based on the corresponding probability evolution description in the master equation has to be the starting point, and its specification for a divergence from the uniform distribution is either a difference of the same formula at U and Q , or not.

As an example, we investigate a power-like nonlinearity, $a(P) = P^\lambda$. In this case, we do not use the Csiszar formula, but solve Equation (19) with the traditional logarithmic ansatz, $f(\xi) = -\ln \xi$:

$$\frac{\partial \sigma}{\partial P_n} = f'(\xi_n) = -\frac{Q_n^\lambda}{P_n^\lambda}. \quad (23)$$

Its particular solution fixing $\rho[Q, Q] = 0$ reads as

$$\rho^{(POW)}[P, Q] = \sum_n Q_n \ln_\lambda \frac{Q_n}{P_n}, \quad (24)$$

with the so called deformed logarithm function [41–43] parametrized with λ :

$$\ln_\lambda x \equiv \frac{1 - x^{\lambda-1}}{1 - \lambda}. \quad (25)$$

We note that starting with an $f(\xi)$ function of a deformed logarithm as above, the result again has the same form, just the new λ is the product of the starting parameter ν in the deformation and the power in the master equation level non-linearity, $\lambda_{\text{new}} = \lambda\nu$. The result Equation (24) is robust. Finally the suggested complexity measure becomes

$$\rho^{(POW)}[U, Q] = \sum_{n=1}^W Q_n \ln_\lambda(WQ_n) = W^{\lambda-1} (S_T[U] - S_T[Q]). \quad (26)$$

It is proportional to a difference of Tsallis entropies,

$$S_T[Q] \equiv - \sum_n Q_n \ln_\lambda Q_n, \quad S_T[U] = -\ln_\lambda(1/W), \quad (27)$$

and hence by construction proves that the Tsallis entropy is also maximal at the uniform distribution, since $\rho \geq 0$ holds, based on the Jensen inequality applied to Equation (24). The forefactor $W^{\lambda-1}$, depending on the number of states, signalizes non-extensivity in this case.

3. LGGR: A Linear Model for Local Growth and Global Resets

The master Equation (3) and its nonlinear generalization (9) describing the evolution of a probability distribution over possible physical microstates are quite general. The familiar diffusion problem in one dimension is described by transition rates between neighboring indices only, allowing

to increase or decrease the index by one in a short time Δt . Their symmetric part is responsible for the diffusion coefficient while their difference provides a driving force for the drift. In higher dimensional diffusion problems there are also some sparse nonzero elements according to the finite size, like $w_{n+N_x, n}$ for a local jump in the positive y direction, etc. These systems in the continuum limit are equivalent to a simple Fokker–Planck equation [44–46].

On the other hand in the up to date research we often face complex systems with underlying dynamics far from the detailed balance: the ratio of reversed micro-transition rates is far from looking as a ratio of a simple given function of indices, $w_{nm}/w_{mn} \neq Q_n/Q_m$. In this paper we present a particular simple model for extremely asymmetric processes: the growth rate is local in the index, while there is a global reset rate from anywhere to the starting (ground) state with index zero. In this local growth, global reset (LGGR) model we assume the following structure for the transition rate matrix ($m \rightarrow n$):

$$w_{nm} = \mu_n \delta_{n,m+1} + \gamma_m \delta_{n,0}. \quad (28)$$

The resulting linear master equation becomes

$$\dot{P}_n = \mu_{n-1} P_{n-1} - (\mu_n + \gamma_n) P_n \quad (29)$$

for $n > 0$. The evolution for the $n = 0$ state, \dot{P}_0 , then includes the summation

$$\dot{P}_0 = \sum_m \gamma_m P_m - (\mu_0 + \gamma_0) P_0, \quad (30)$$

but alternatively P_0 can be obtained from the normalization condition, $P_0 = 1 - \sum_{n=1}^{\infty} P_n$. Now the complexity measure is given by

$$\rho^{(LGGR)}[U, Q] = \sum_{n=1}^W Q_n f^{(KL)}(W Q_n) = \ln W + \sum_{n=1}^W Q_n \ln Q_n, \quad (31)$$

and we are especially interested in its value for the stationary distribution of a given LGGR model. So the missing link is the computation of the stationary PDF in terms of μ_n and γ_n .

Here we review some of the most important LGGR models. The stationary distribution is obtained by the recursive resolution of the $\dot{Q}_n = 0$ condition,

$$Q_n = \frac{\mu_{n-1}}{\mu_n + \gamma_n} Q_{n-1} = Q_0 \prod_{j=1}^n \frac{\mu_{j-1}}{\mu_j + \gamma_j}. \quad (32)$$

For constant growth and reset rates, $\mu_n = \sigma$, $\gamma_n = \gamma$, the result is the exponential distribution,

$$Q_n = Q_0 e^{-\beta n} \quad (33)$$

with $\beta = \ln(1 + \gamma/\sigma)$. In this LGGR model with the Kullback–Leibler divergence and the traditional Boltzmannian entropy formula also non-exponential stationary distributions emerge if the transition rates are state-dependent. With a growth rate featuring linear preference, $\mu_n = \sigma(n + b)$, the stationary solution is a Waring distribution [47],

$$Q_n = Q_0 \frac{(b)_n}{(c)_n}, \quad (34)$$

using the Pochhammer symbol $(a)_n = a(a+1) \dots (a+n-1)$ and $c = b+1 + \gamma/\sigma$. It exemplifies a power-law decay for large n , $Q_n \sim n^{-1-\gamma/\sigma}$.

In the continuous LGGR version we consider the evolution equation

$$\frac{\partial}{\partial t} \mathcal{P}(x, t) = -\frac{\partial}{\partial x} (\mu(x) \mathcal{P}(x, t)) - \gamma(x) \mathcal{P}(x, t) + A(t) \delta(x) \quad (35)$$

with

$$A(t) \equiv \int_0^{\infty} \gamma(x) \mathcal{P}(x, t) dx - \mu(0) \mathcal{P}(0, t). \quad (36)$$

This quantity may vanish for some special cases where $\mu(0) = 0$ and $\int_0^{\infty} \gamma(x) \mathcal{P}(x, t) dx = 0$. The corresponding stationary distribution becomes

$$Q(x) = \frac{\mu(0)Q(0)}{\mu(x)} \exp \left(- \int_0^x \frac{\gamma(u)}{\mu(u)} du \right). \quad (37)$$

In particular the stationary PDF belonging to the linear local growth rate, $\mu(x) = \sigma(x + b)$, featuring linear preference and combined with a constant reset rate, $\gamma(x) = \gamma$, is given by the Tsallis–Pareto distribution:

$$Q(x) = \frac{\gamma}{\sigma b} \left(1 + \frac{x}{b} \right)^{-1-\gamma/\sigma}. \quad (38)$$

So power-law tailed distributions are stationary solutions to linear probability evolution equations with the traditional entropy and entropic divergence formula, if a constant global reset rate is connected with a linear preference in the local growth rate.

The total balance condition (37) determining the stationary PDF, on the other hand, can be viewed as a generalized fluctuation–dissipation theorem [48]. This form is useful for cases when the stationary PDF, $Q(x)$, can be easily observed: than one of the rates provides information on the other. Simple integration of Equation (35) with $\dot{Q}(x) = 0$ leads to such a relation, expressing the local growth rate with a distribution-tail average of the reset rate:

$$\mu(x) = \frac{1}{Q(x)} \int_x^{\infty} \gamma(u) Q(u) du. \quad (39)$$

For an exponential PDF, $Q(x) = e^{-\beta x}/Z$, and constant reset rate, $\gamma(x) = \gamma$, one obtains $\mu = \gamma/\beta$ in analogy to the relation between diffusion rate and dissipation coefficient [49,50]. In general, for constant global reset rate γ , $\mu(x)/\gamma$ becomes the inverse hazard rate [51]. Such features transform LGGR to a simple and effective candidate for being an “Ising model”. It can be applied to far from equilibrium, with no detailed balance, and with complex system dynamics with various transition rates describing preferential scenarios.

4. Citation Popularity, Income Distribution and Some Further Application Areas

Finally we would like to review a few applications of this simplified LGGR treatment. As it has been published in [26], stationary PDF-s can be reproduced in several socio-economic or physical models. Socio-economic examples include popularity ranking from citation numbers [52–54], income [55–57] and settlement sizes [58]. For biological and ecological systems population abundance [59] on different levels (species, spatial, etc.) can be studied. For physical systems a good example is the hadron number distribution in high energy accelerator collision events [60].

In exceptional cases the income data are known on the level of individuals and are available for an extended period over several years. In such cases they allow for testing the growth and the entry/exit rates experimentally. We emphasize here that simply observing a certain distribution, $Q(x)$, does not decide whether it is stationary, and even less whether an LGGR model could satisfactorily describe the real dynamics in its background. Whenever both observed distributions and transition rates with unidirected small-step growth and big-step global resets can be observed, then the LGGR can be verified or falsified.

As already discussed in earlier publications, citation popularity data follow a continuous Tsallis–Pareto distribution [52] which, scaled with the average number of citations, contains only one nontrivial fit parameter. In our analysis we have included Web of Science citations of some journals and institutes with international reputation as well as individuals. Furthermore Facebook likes and shares were counted, and similarly Youtube likes were recorded for diverse web pages. Surprisingly, all these popularity data collapse onto the same universal curve. The observed distributions can be understood by assuming linear preference in the local growth rate and a constant reset rate in the continuous version of the dynamics [52].

It is methodically useful to report here briefly about our recent investigations on income distribution. The starting point for our study is an individual level, complete income data set for a nine year period (2001–2009) in Cluj county (Romania) [61]. Beside the observed $Q(x)$ PDF for income, we can follow the individual level income data to justify assumptions on the rates $\mu(x)$ and $\gamma(x)$. Our preliminary findings include an improved ansatz on these rates. The reset rate, $\gamma(x)$, occurs to be intelligent: at low income level it is negative, people enter into the income system mainly at this level, while at high income it is positive, the income is naturally high when they are close to retirement. The following fit describes well the averaged data

$$\gamma(x) = K - \frac{b}{x + q}, \quad (40)$$

with K , b and q freely adjustable parameters.

The local growth rate, $\mu(x)$, contains a simple linear term since the income (salary) increase is as a rule in per cent, so the amount of increase is proportional to the income at that moment. The assumed

$$\mu(x) = \beta x \quad (41)$$

functional form fits well the income data in Romania, Cluj district in the years 2001–2009.

For these growth and reset rates the resulting PDF is analytically obtained: using Equation (29), and normalizing it we obtain the Beta Prime distribution

$$Q(x) = q^{\left(\frac{K}{\beta} - \frac{b}{\beta q}\right)} \frac{\Gamma\left(\frac{b}{q\beta}\right)}{\Gamma\left(\frac{b}{q\beta} - \frac{K}{\beta}\right) \Gamma\left(\frac{K}{\beta}\right)} \left(1 + \frac{x}{q}\right)^{-\frac{b}{\beta q}} x^{\frac{b}{\beta q} - \frac{K}{\beta} - 1}. \quad (42)$$

This formula, normalized and expressed in terms of $z = x/\langle x \rangle$ contains only one parameter, and income distributions from Romania, Hungary, USA, Finland and Australia fall onto the common universal curve, $\langle x \rangle Q(x) = 12z(1+z)^{-5}$.

Finally, let us mention the hadronization process in high energy elementary particle and heavy-ion collisions as a possible application field. Here, two distributions seem not to have a microscopic dynamical explanation, namely the multiplicity distribution of hadron number in a series of collisions with the same total energy, and the one-particle kinetic energy distribution averages over such collision events. The former seems to follow a negative binomial distribution (NBD) [62], although at the Large Hadron Collider (LHC) energies this is just a first approximation: the real distribution, including higher and higher kinetic energies in the count might require more sophisticated descriptions [63]. The latter on the other hand is beautifully fitted by a Tsallis–Pareto distribution in the mid-rapidity range, best interpolating between low, mid and higher transverse momenta in the range $p_T = 0.8 \dots 12$ GeV [60,64].

We note that these distributions can be related to each other: a simple filling of a phase space volume whose dimensionality changes with the fluctuating number of produced hadrons event by event relates the NBD and the Tsallis–Pareto power-law tailed distribution in the individual kinetic energy. It is easy to demonstrate as follows.

In kinetic models, the microcanonical phase space is an energy shell in high dimension. The constant total energy constraint is a norm on the individual momenta, in a one-dimensional extreme relativistic jet this norm is L_1 , $E = \sum_{i=1}^n |p_i|$, in a two-dimensional section at mid-rapidity of massive, non-relativistic particles on the other hand it is Pythagorean, $E = \sum_{i=1}^{2n} p_i^2 / 2m$. Both formulas are given for n particles. In a general asset we are interested in the surface of an N -ball with radius R in L_p norm, which is known to be [65]

$$V_N^{(p)}(R) = \frac{[2R\Gamma(1+1/p)]^N}{\Gamma(1+N/p)}. \quad (43)$$

One specifies $N = n, p = 1$ and $R(E) = E$ for the jets and $N = 2n, p = 2$ with $R(E) = \sqrt{2mE}$ for the two-dimensional ideal non-relativistic gas. The calculations deliver the results $V_n^{(1)}(E) = (2E)^n / n!$ and $V_{2n}^{(2)}(\sqrt{2mE}) = (\pi m E)^n / n!$. The microcanonical constraint volumes are obtained by a simple derivation from these formulas, $\Omega_N^{(p)}(E) = dV_N^{(p)}(R(E)) / dE$. Finally the phase space ratio (conditional probability) for a selected single particle having ε energy from this total of E in both cases is given by

$$r_n^{(1)}(\varepsilon, E) = r_{2n}^{(2)}(\varepsilon, E) = \frac{\Omega(E - \varepsilon)}{\Omega(E)} = \frac{n-1}{E} \left(1 - \frac{\varepsilon}{E}\right)^{n-2}. \quad (44)$$

This ratio is to be averaged over an NBD distribution of the newly produced $n - 2$ hadrons (2 is minimally needed for making the collision),

$$Q_n = \binom{n+k}{n} f^n (1+f)^{-n-k-1}. \quad (45)$$

The result is a Tsallis–Pareto distribution [66] with some modifying factors:

$$Q(\varepsilon) = \frac{1}{E} \left[1 + f(k+1) - f k \frac{\varepsilon}{E}\right] \left(1 + f \frac{\varepsilon}{E}\right)^{-k-2}. \quad (46)$$

For large $\langle n \rangle \gg 1$, characteristic for central heavy ion collisions, it approaches the classical form

$$Q(\varepsilon) \approx \frac{1}{T} \left(1 + \frac{1}{k+1} \frac{\varepsilon}{T}\right)^{-k-2}, \quad (47)$$

with $T = E / \langle n \rangle = f(k+1)$, while for small $\langle n \rangle \ll 1$, more proper for describing elementary pp collisions, a similar Tsallis–Pareto law emerges with a modified power:

$$Q(\varepsilon) \approx \frac{1}{E} \left(1 + \frac{1}{k+1} \frac{\varepsilon}{T}\right)^{-k-1}. \quad (48)$$

Equation (46) represents a monotonic falling distribution in ε in the interval $[0, E]$ and its integral over this interval is normalized to 1. Several measurements agree with this or similar functional forms, although some details can also be fitted with different forefactors to the Tsallis–Pareto term [67].

5. Summary

In conclusion, we presented an H-theorem like proof for stochastic evolution without detailed balance, in particular we have shown that Csizsár's f-divergence formula for the entropic divergence shrinks between two arbitrary starting probability distributions for $f'' > 0$. The condition for this statement is that the same master equation, linear in the probability, governs the evolution of both. Further we presented the local growth global reset (LGGR) model, as a simple approach to processes

far from the detailed balance and yet leading to interesting stationary distributions. This model utilizes only two rates, $\mu(x)$ for the growth and $\gamma(x)$ for the reset to the $x = 0$ point in the state space. A fluctuation-dissipation relation connects these rates via the normalized tail probability [68] of the stationary PDF, cf. Equation (39). Application fields for LGGR type approaches are numerous, here we exemplified the citation popularity, the income distribution and the hadron distribution in high energy experiments.

Author Contributions: Conceptualization and methodology was performed by T.S.B.; Formal analysis, data validation and discussion was shared by all authors; Conference talk and draft preparation as well as funding acquisition for this, including APC was provided by T.S.B.

Funding: This research and the APC for this paper was funded by the Hungarian National Research, Development and Innovation Office, grant number K123815. Further support was achieved from the UEFISCDI research grant PN-III-P4-ID-PCCF-2016-0084.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marton, K. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **1996**, *6*, 556–571. [CrossRef]
2. Talagrand, M. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **1996**, *6*, 587–600. [CrossRef]
3. Villani, C. *Optimal Transport: Old and New*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 338.
4. Guillin, A.; Léonard, C.; Wu, L.; Yao, N. Transportation-information inequalities for Markov processes. *Probab. Theory Relat. Field* **2009**, *144*, 669–695. [CrossRef]
5. Kullback, S. *Information Theory and Statistics*; Courier Corporation: Paris, France, 1997.
6. Fathi, M.; Shu, Y. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli* **2018**, *24*, 672–698. [CrossRef]
7. Cheng, L.; Li, R.; Wu, L. Ricci curvature and W_1 -exponential convergence of Markov processes on graphs. *arXiv* **2019**, arXiv:1907.11036.
8. Boltzmann, L. *Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmeleichgewicht*; Kk Hof- und Staatsdruckerei: Wien, Austria, 1877.
9. Gibbs, J.W. *Elementary Principles in Statistical Mechanics*; Yale University Press: New Haven, CT, USA, 2014.
10. Planck, M. *Über den zweiten Hauptsatz der mechanischen Wärmetheorie*. Inaugural Dissertation zur Erlangung der Philosophischen Doktorwürde, K. Universität München, 1879, Theodor Ackermann, München, 1879. Available online: <https://edoc.hu-berlin.de/bitstream/handle/18452/734/planck.pdf?sequence=1> (accessed on 9 October 2019).
11. Planck, M. Entropy and Temperature of Radiant Heat. *Ann. Phys.-Berlin* **1900**, *1*, 719–737. [CrossRef]
12. Shannon, C.E. A mathematical theory of communication. *Bell Labs Tech. J.* **1948**, *27*, 379–423. [CrossRef]
13. Rényi, A. On the dimension and entropy of probability distributions. *Acta Math. Hung.* **1959**, *10*, 193–215. [CrossRef]
14. Rényi, A. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol 1: Contributions to the Theory of Statistics*, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Oakland, CA, USA, 1961; p. 767.
15. Aczél, J.; Daróczy, Z. On measures of information and their characterizations. In *Mathematics in Science and Engineering*; Academic Press: New York, NY, USA, 1975; Volume 115, p. 168.
16. Aczel, J.; Forte, B. Generalized entropies and the maximum entropy principle. In *Maximum Entropy and Bayesian Methods in Applied Statistics*; Cambridge University Press: Cambridge, UK, 1986; pp. 95–100.
17. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
18. Deng, Y. Deng entropy. *Chaos Solitons Fractals* **2016**, *91*, 549–553. [CrossRef]
19. Khintchine, A. Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **1934**, *109*, 604–615. [CrossRef]
20. Khinchin, A.Y. The concept of entropy in the theory of probability. *Uspekhi Mat. Nauk* **1953**, *8*, 3–20.

21. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: Berlin/Heidelberg, Germany, 2009.
22. Song, Y.; Deng, Y. Divergence Measure of Belief Function and Its Application in Data Fusion. *IEEE Access* **2019**, *7*, 107465–107472. [[CrossRef](#)]
23. Song, Y.; Deng, Y. A new method to measure the divergence in evidential sensor data fusion. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1550147719841295. [[CrossRef](#)]
24. Fei, L.; Deng, Y. A new divergence measure for basic probability assignment and its applications in extremely uncertain environments. *Int. J. Intell. Syst.* **2019**, *34*, 584–600. [[CrossRef](#)]
25. Aczél, J.; Forte, B.; Ng, C.T. Why the Shannon and Hartley entropies are ‘natural’. *Adv. Appl. Probab.* **1974**, *6*, 131–146. [[CrossRef](#)]
26. Biró, T.S.; Nédá, Z. Unidirectional random growth with resetting. *Physica A* **2018**, *499*, 335–361. [[CrossRef](#)]
27. Chen, A.; Lu, Y.; Ng, K.; Zhang, H. Markov branching processes with killing and resurrection. *Sci. China-Math.* **2016**, *59*, 573–588. [[CrossRef](#)]
28. Biró, T.; Telcs, A.; Nédá, Z. Entropic Distance for Nonlinear Master Equation. *Universe* **2018**, *4*, 10. [[CrossRef](#)]
29. Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255. [[CrossRef](#)]
30. Csiszár, I. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **1975**, 146–158. [[CrossRef](#)]
31. Csiszár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kut. Int. Közl.* **1963**, *8*, 85–108.
32. Jensen, J.L.W.V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **1906**, *30*, 175–193. [[CrossRef](#)]
33. Lieb, E.H. Some convexity and subadditivity properties of entropy. In *Inequalities*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 67–79.
34. Lagrange, J.L. *Théorie des Fonctions Analytiques*; Ve. Courcier: Paris, France, 1797.
35. Gorban, A.N.; Gorban, P.A.; Judge, G. Entropy: The Markov Ordering Approach. *Entropy* **2010**, *12*, 1145–1193. [[CrossRef](#)]
36. Zhou, Y.; Yin, C. Power-law Fokker-Planck equation of unimolecular reaction based on the approximation to master equation. *Physica A* **2016**, *463*, 445–451. [[CrossRef](#)]
37. Boon, J.P.; Lutsko, J.F.; Lutsko, C. Microscopic approach to nonlinear reaction-diffusion: The case of morphogen gradient formation. *Phys. Rev. E* **2012**, *85*, 021126. [[CrossRef](#)]
38. Borland, L. Microscopic dynamics of the nonlinear Fokker-Planck equation: A phenomenological model. *Phys. Rev. E* **1998**, *57*, 6634–6642. [[CrossRef](#)]
39. Bertsch, G.F.; Gupta, S.D. A guide to microscopic models for intermediate energy heavy ion collisions. *Phys. Rep.* **1988**, *160*, 189–233. [[CrossRef](#)]
40. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
41. Naudts, J. Deformed exponentials and logarithms in generalized thermostatistics. *Physica A* **2002**, *316*, 323–334. [[CrossRef](#)]
42. Borges, E.P. A possible deformed algebra and calculus inspired in nonextensive thermostatistics. *Physica A* **2004**, *340*, 95–101. [[CrossRef](#)]
43. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Two-parameter deformations of logarithm, exponential, and entropy: A consistent framework for generalized statistical mechanics. *Phys. Rev. E* **2005**, *71*, 046128. [[CrossRef](#)] [[PubMed](#)]
44. Planck, M. Fokker-Planck equation. *Sitzungsber. Preuß. Akad. Wiss* **1917**, *3*, 324–341.
45. Fokker, A.D. Ein invarianter Variationssatz für die Bewegung mehrerer elektrischer Massenteilchen. *Z. Phys.* **1929**, *58*, 386–393. [[CrossRef](#)]
46. Risken, H. *The Fokker-Planck Equation*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 63–95.
47. Irwin, J.O. The generalized Waring distribution applied to accident theory. *J. R. Stat. Soc. Ser. A-Stat. Soc.* **1968**, *131*, 205–225. [[CrossRef](#)]
48. Bochkov, G.N.; Kuzovlev, Y.E. Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics: I. generalized fluctuation-dissipation theorem. *Physica A* **1981**, *106*, 443–479.
49. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys. (Berl.)* **1905**, *322*, 549–560. [[CrossRef](#)]

50. Einstein, A. Zur Theorie der Brownschen Bewegung. *Ann. Phys. (Berl.)* **1906**, *324*, 371–381. [[CrossRef](#)]
51. Rausand, M.; Hoyland, A. *System Reliability Theory: Models, Statistical Methods, and Applications*; John Wiley & Sons: Chichester, UK, 2004.
52. Nédá, Z.; Varga, L.; Biró, T.S. Science and Facebook: The same popularity law ! *PLoS ONE* **2017**, *12*, e0179656. [[CrossRef](#)]
53. Schubert, A.; Glänzel, W. A dynamic look at a class of skew distributions. A model with scientometric applications. *Scientometrics* **1984**, *6*, 149. [[CrossRef](#)]
54. Glänzel, W.; Schubert, A. Predictive aspects of a stochastic model for citation processes. *Inf. Process. Manag.* **1995**, *31*, 69. [[CrossRef](#)]
55. Chakraborti, A.; Chatterjee, A.; Chakrabarti, B.; Chakravarty, S.R. *Econophysics of Income and Wealth Distributions*; Cambridge University Press: Cambridge, UK, 2013.
56. Piketty, T.; Goldhammer, A. *Capital in the Twenty-First Century*; The Belknap Press of Harvard University Press, Cambridge, MA, USA, 2014.
57. Bouchaud, J.P.; Mézard, M. Wealth condensation in a simple model of economy. *Physica A* **2000**, *282*, 536–545. [[CrossRef](#)]
58. Bee, M.; Riccaboni, M.; Schiavo, S. Distribution of City Size: Gibrat, Pareto, Zipf. In *The Mathematics of Urban Morphology*; Birkhäuser: Cham, Switzerland, 2019; pp. 77–91.
59. Horvát, S.; Derzsi, A.; Nédá, Z.; Balog, A. A spatially explicit model for tropical tree diversity patterns. *J. Theor. Biol.* **2010**, *265*, 517–523. [[CrossRef](#)] [[PubMed](#)]
60. Bíró, G.; Barnaföldi, G.G.; Papp, G.; Biró, T.S. Multiplicity Dependence in the Non-Extensive Hadronization Model Calculated by the HIJING Framework++. *Universe* **2019**, *5*, 134. [[CrossRef](#)]
61. Derzsy, N.; Neda, Z.; Santos, M.A. Income distribution patterns from a complete social security database. *Physica A* **2012**, *391*, 5611–5619. [[CrossRef](#)]
62. Fisher, R.A. The negative binomial distribution. *Ann. Eugen.* **1941**, *11*, 182–187. [[CrossRef](#)]
63. Grosse-Oettringshaus, J.I.; Rygers, K. Charged-Particle Multiplicity in Proton-Proton Collisions. *J. Phys. G* **2010**, *37*, 083001. [[CrossRef](#)]
64. Bíró, G. Application of New-Generation High Energy Physical Detector Simulators to the Investigation of Identified Hadron Spectra. MSc Thesis, Eötvös University, Budapest, Hungary, 2015.
65. Barthe, F.; Guédon, O.; Mendelson, S.; Naor, A. A probabilistic approach to the geometry of the ℓ_p n-ball. *Ann. Probab.* **2005**, *33*, 480–513. [[CrossRef](#)]
66. Pareto, V. *Manuale di Economia Politica*; Societa Editrice: Torino, Italy, 1906; Volume 13.
67. Bíró, G.; Barnaföldi, G.G.; Biró, T.S.; Ürmösy, K.; Takács, A. Systematic Analysis of the Non-Extensive Statistical Approach in High Energy Particle Collisions-Experiments vs Theory. *Entropy* **2017**, *19*, 88. [[CrossRef](#)]
68. Telcs, A.; Glänzel, W.; Schubert, A. Characterization and statistical test using truncated expectations for a class of skew distributions. *Math. Soc. Sci.* **1985**, *10*, 169. [[CrossRef](#)]

