Eur. Phys. J. B **49**, 491–511 (2006)
DOI: 10.1140/epjb/e2006-00084-2

THE EUROPEAN PHYSICAL JOURNAL B

The significant digit law: a paradigm of statistical scale symmetries

The significant digit law

A. Pocheau[a]

IRPHE, CNRS & Universités Aix-Marseille I & II, 49 rue Joliot-Curie, B.P. 146, Technopôle de Château-Gombert, 13384 Marseille, Cedex 13, France

Received 5 September 2005
Published online 31 March 2006 –

Abstract. In many different topics, the most significant digits of data series display a non-uniform distribution which points to an equiprobability of logarithms. This surprising ubiquitous property, known as the significant digit law, is shown here to follow from two similar, albeit different, scale symmetries: the *scale*-invariance *and* the *scale-ratio* invariance. After having legitimized these symmetries in the present context, the corresponding symmetric distributions are determined by implementing a *covariance* criterion. The logarithmic distribution is identified as the only distribution satisfying *both* symmetries. Attraction of other distributions to this most symmetric distribution by dilation, stretching and merging is investigated and clarified. The natures of both the scale-invariance and the scale-ratio invariance are further analyzed by determining the structure of the sets composed by the corresponding symmetric distributions. Altogether, these results provide new insights into the meaning and the role of scale symmetries in statistics.

PACS. 02.50.-r Probability theory, stochastic processes, and statistics – 07.05.Kf Data analysis: algorithms and implementation; data management – 05.40.-a Fluctuation phenomena, random processes, noise, and Brownian motion

1 Introduction

More than a century ago, the astronomer Simon Newcomb noticed a peculiar statistical feature in the distributions of many numerical data sets [1]:

The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally likely.

A consequence of this statistical property is that the probability $P(s)$ of occurrence of the first significant digit s in many common data sets is actually non-uniform: in contrast with basic intuition, the smaller digits (e.g. [1]) appear more frequently than the larger digits (e.g. [9]) (see Tab. 1). Moreover, this non-uniform occurrence follows a definite distribution whose shape is logarithmic (Fig. 1a). This is known as the "significant-digit law", hereafter labelled "SDL". In the decimal system, it reads [1–3]:

$$P(s) = \log_{10}(1 + s^{-1}). \qquad (1)$$

An amazing empirical support to this logarithmic distribution was put forward by S. Newcomb who noticed that the first pages of logarithmic tables are more soiled — and therefore more visited — than the last pages [1]! Later on, Franck Benford, a physicist of the General Electric Company, provided a strong quantitative support to this law by analysing more than twenty data sets referring to topics as different as areas of rivers, front pages of newspapers, base-ball statistics, street addresses in human groups ... [2] (see Tab. 2 and Fig. 1b).

To emphasize the wide domain of validity of this law and its large universality, we have reported in Figure 2a the histograms of the most significant digits of the *measurements* reported in the articles of a previous issue of this journal [4]. We have complemented it in Figure 2b by the histograms of the most significant digits of the *first page* of the articles to which they refer. These data, of course, refer to very different topics and, within each, to mostly uncorrelated features. Their normalized histograms are however close to the SDL for each issue of the journal and get even closer to it when gathered in a common data set.

Nowadays, the significant-digit law is used in computer design for saving memory [5,6], in tax control as a mean for detecting frauds [7], in mathematical modeling and

[a] e-mail: `alain.pocheau@irphe.univ-mrs.fr`

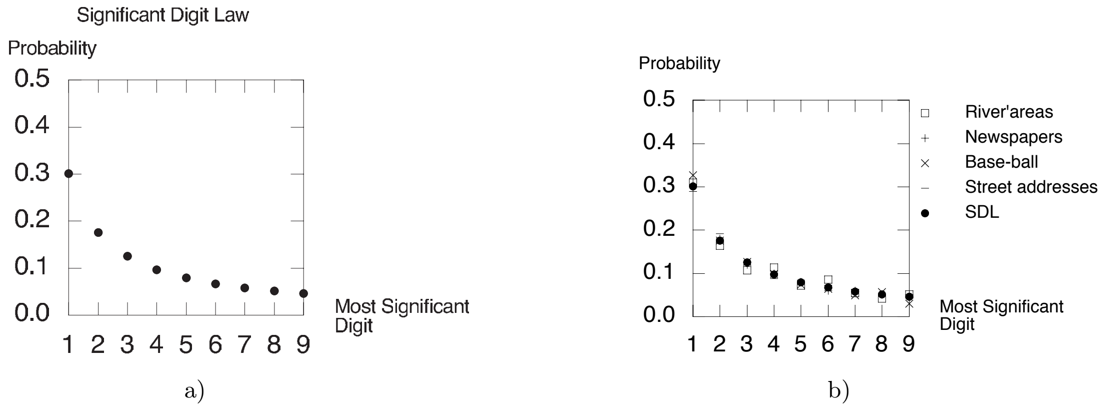

Fig. 1. Significant digit law (SDL) and Bendford's evidences. (a) Probability of occurrence of the most significant digit in common data sets according to the SDL (Tab. 1) (1). (b) Normalized histograms of the most significant digit of data corresponding to areas of rivers, numbers displayed on front pages of newspapers, base-ball statistics, street addresses in human groups . . . [2] (Tab. 2).

Table 1. Probability of occurrence of the most significant digit according to the SDL (1)

digit	1	2	3	4	5	6	7	8	9
probability	0.301	0.176	0.125	0.097	0.080	0.067	0.058	0.0518	0.046

Table 2. Some of the distributions of most significant digits obtained by Franck Bendford on topics as different as areas of rivers, numbers displayed on front pages of newspapers, base-ball statistics, street addresses in human groups . . . [2]. Probabilities are given in percents.

Digit	1	2	3	4	5	6	7	8	9	Sample
River's areas	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Newpapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Base-ball	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Street addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342

in finance [10]. Yet, its origin and its significance still remain the subject of mathematical [3,11–13] and statistical [14,15] investigations. The latter studies [14,15] have aimed at pointing out the statistical processes or the stochastic systems that can give rise to the significant-digit law; the former studies [3,11–13] have looked at identifying the statistical properties that are necessary for exhibiting this peculiar law. Among them, those based on the concepts of scale-invariance [3,11] seem both the most promising and the most amenable to extension to other fields.

However, two pitfalls may be encountered when applying scale symmetries in statistics: the possible breaking of a probability measure depending on the set on which distributions are considered (Sect. 3.1.1 and App. A.5); the variation of the normalization constants of distributions in comparing the effects of scale change on them (App. A.1). To avoid them, we shall work with conditional probabilities defined on bounded intervals (Sect. 3.1.2). This framework, different from those used in previous studies [3,11], will yield us to draw different conclusions on this topic.

The present paper aims at showing how the properties of *scale* invariance *and* of *scale-ratio* invariance, when applied *altogether* on sets involving a *non-degenerate* probability measure, succeed in selecting a universal attractor distribution: the significant digit law.

For this, we begin in Section 2 by justifying the relevance of these statistical invariances to large enough or various enough data sets. In Section 3, we implement from these symmetries the frequency distribution of data on *bounded*, strictly positive, semi-closed intervals. Two different classes of distributions referring either to scale invariance or to scale-ratio invariance are found. Interestingly, the significant digit law is recognized in Section 4 as their *common* distribution. Its attraction on distributions by dilation, stretch and merge is then investigated and clarified.

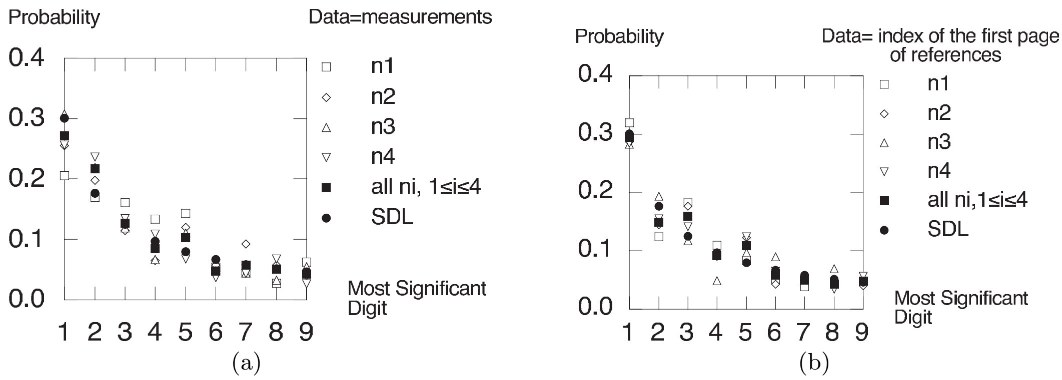

Fig. 2. Example of the SDL and of the convergence by merging to the SDL. Data are taken from the Vol. 22 of Eur. Phys. J. B [4]. They are collected in four sets corresponding to the numbers $n_i = 1, 2, 3, 4$ of this volume and finally gathered in a common data set. The normalized histograms of the most significant digit of these different data sets are then compared to the SDL. Each set $n_i = 1, 2, 3, 4$ follows the general trend of the SDL. However, their common data set gets even closer to it: the more data, the closer to the SDL in average. (a) Data correspond to the measurements reported in articles. (b) Data correspond to the first page of the cited articles.

This selection of a universal attractor distribution by symmetries indicates that the present statistical problem is appropriate for investigating the nature of scale symmetries and for providing an insight into their interplay. To this goal, we address in Section 5 from a general viewpoint the concepts of characteristic scale and of characteristic scale-ratios for distributions and we determine, for each symmetry, the structure of the corresponding sets of symmetric distributions.

A conclusion about the significance of the link between the significant digit law and the scale/scale-ratio symmetries is finally given in Section 6.

2 Scale invariance; scale-ratio invariance

2.1 Significant digit law and symmetries

Whereas Newcomb's law may be satisfied on some data sets, it largely fails when data are correlated with some deterministic feature. This is the case for instance for telephone numbers since some of their digits correspond to the domestic/abroad calling direction, the state or the region. Another example is given by personal identification numbers since a part of them refers to the sex number, the birth year or the birth place. There, digits are set by definite features that privilege some numbers, even up to forbidding some others. Clearly, the logarithmic distribution cannot hold in these instances.

Conversely, for data sets satisfying Newcomb's law, the above remark calls for questioning the existence of *any* noticeable feature in their distribution. The guess that there is *none* is supported by the following considerations:

1. Uncorrelation of events
   The data sets that are known to satisfy Newcomb's law refer to events which seem to be uncorrelated. Examples include the share volumes on stock exchange of various products [7], the income tax from many different companies and products [7,8], the stock market indexes [9], the numbers printed on front pages of newspapers [2,12,17], the street address among a large group of scientists [2], the radioactive half-lives of many different elements [18] . . . Here, uncorrelation of events supports the *absence* of characteristic features in the corresponding data files.
2. Convergence by merging
   Merging together data sets not satisfying Newcomb's law yields a distribution closer to Newcomb's law than the initial ones [10] (Fig. 2). Here, merging implies two important features for the new data file: the increase of the number of data; the gathering of the features brought about by each data set. Increase of data number indeed enhances the convergence of normalized histograms to a limit distribution. However, by itself, this cannot be taken as responsible for an evolution of the limit distribution. On the other hand, merging different features into a common data set makes each of them less discernible in the average: the resulting file then actually better approaches a feature-*free* data series.
3. Convergence by stretching
   When a data file does not initially follow Newcomb's law, raising it to higher and higher powers makes it approach a logarithmic distribution [13] (Figs. 5 and 6). As raising data to powers stretches their distribution, a possibility is that *any* characteristic feature gets so "diluted" among the file that it is finally "wiped out".
4. Stability by homothety
   When a data file initially follows Newcomb's law, any of its homothetic files still satisfies a logarithmic distribution. In the same spirit, we notice that, if data series involve no specific intrinsic features, so do any homothetic series. The robustness of both these properties to homothety shows the compatibility of Newcomb's law with the *absence* of intrinsic feature in data series.
5. Universality
   The universal character of the significant digit law indicates that it refers to a widely shared property. Accordingly, one may expect that this property is rather

linked to the *absence* of specificities rather than to the existence of some definite one.

Following the above considerations, Newcomb's law may be expected to rely on deep symmetries. Good candidates are those symmetries which are related to the concept of scale. We consider two of them below: scale-invariance and scale-ratio invariance.

2.2 Scale invariance

2.2.1 Relevance

A paramount concept in the study of phenomena is that of "magnitude". It is usually referred to as the concept of "scale" since it actually reduces to it in a geometrical context.

When a definite magnitude is singled out by the process or by the state under study, the corresponding scale is definitely particularized within the resulting data series, even in a statistical framework. It then stands as one of their characteristic features. However, following Section 2.1, one may guess that, although the concept of characteristic scale is relevant to most of usual data series, it may be incompatible with some. To better realize this property, let us address the implications of dilation and merge to the present issue and question the nature of the limit distribution.

We first notice that, given a phenomenon exhibiting a magnitude l, the particular scale that it drives in data series is the *number* $n_l = l/l_s$ equal to the ratio of the magnitude l to the magnitude l_s of the standard used to measure it. However, different observers using different standards l_s would conclude to different characteristic scales n_l for the data series representative of the phenomenon. In particular, observers using homothetic standards would conclude to homothetic series involving homothetic characteristic scales. Then, comparing these series would give the same impression as if the standard l_s had been kept fixed but the characteristic magnitudes l had been dilated. Merging such data series would thus result in scattering characteristic scales among data: merging dilated files tends to destroy the concept of characteristic scale.

Pushing this property at the limit, we may obtain, by successive change of phenomena or of standard units and by merging, a file containing *no* characteristic scale n_l. This property implies that the resulting data refer to a set of phenomena *altogether* showing *no* characteristic magnitude l, so that all standards l_s are thus *equivalent* regarding their measurements. Stated differently, whereas changing a standard l_s for another l'_s actually induces a dilation of data, i.e. a scale dilation, this nevertheless must *not* change the statistical representation of the file: there must be statistical *scale-invariance.*

2.2.2 Statistical equivalence; scale covariance

In the above analysis, merging is considered to play a role similar as that of thermalization in statistical mechanics. However, in the same way as thermalization does not mean uniform distribution in phase space, scale-invariance is not synonymous of uniform probability. In particular, the shapes of scale-invariant distributions may be identified from the symmetry which underlies scale-invariance: the invariance by dilation of their representation.

Let us call D_λ the dilation by scale factor λ:

$$D_\lambda(.) : x \to \lambda x = D_\lambda(x). \tag{2}$$

Scale invariance states that dilating *all* scales by the *same* factor λ keeps the statistical representation *unchanged.* This means that homothetic data files are *statistically equivalent* so that their distributions are the *same.* In other words, scale dilation fails in changing the distribution functions: they are scale-*covariant* by dilation D_λ. Reciprocally, any distribution satisfying this property would not allow, by itself, *any* distinction between scales to be made. It could therefore only refer to data files exhibiting *no* characteristic scale, i.e. to scale-invariance. Statistical scale-invariance is thus equivalent to covariance of distributions by scale dilation.

The major interest of covariance by dilation is to provide us with a definite operational criterion for identifying the analytic form of scale-invariant distributions. This criterion will be implemented in Section 3.2.

2.3 Scale-ratio invariance

2.3.1 Relevance

Whereas the significant digit law is expected to satisfy scale-invariance, we have no certainty that the converse is true: some scale-invariant distributions may not satisfy Newcomb's law. If that was the case, Newcomb's law might refer to *more* symmetries than just scale-invariance. Anticipating on this possibility, our problematic consists in determining the possible candidates for such additional symmetries and in identifying the shape of their distributions.

Having adopted scale-invariance as a fundamental symmetry underlying Newcomb's law, an important constraint restricts the range of possible additional symmetries involved in this law: their compatibility with scale-invariance. Equivalently, this turns out forbidding the generation of *any* characteristic scale by the transformations linked to additional symmetries. On the opposite indeed, applying these symmetries to scale-invariant files satisfying Newcomb's law would generate files still satisfying this law while involving some characteristic scales, in contradiction with our basic assumption.

We thus forbid the generation of *any* characteristic scale by the transformations linked to the sought additional symmetries. As these transformations are *single* variate functions, they can therefore only be power laws $S_{x_f,\nu}$:

$$S_{x_f,\nu}(.) : \forall x\ ,\ x \to x_f \left(\frac{x}{x_f}\right)^\nu = S_{x_f,\nu}(x). \tag{3}$$

Here ν parametrizes the stretch induced by the transformation whereas x_f denotes its non-zero fixed point.

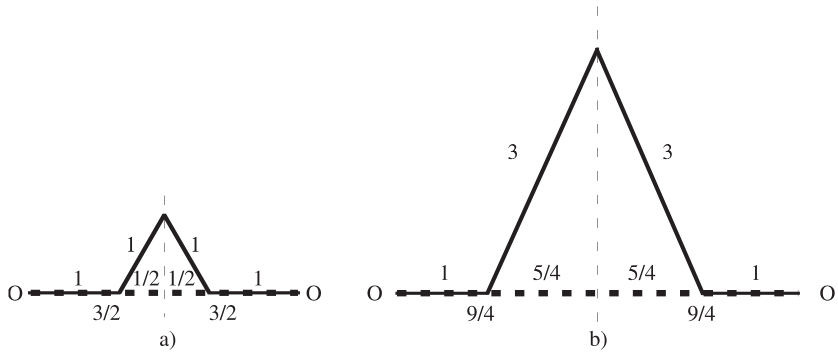

Fig. 3. Sketch of the effect of a stretching transformation $S_{x_f,\nu}(x) = x_f(x/x_f)^\nu$ on an elementary pattern (a) of a Koch curve. Stretching is applied in (b) from each of the extremities O to the symmetry axis, on both the mother branch (the horizontal line) and the daughter branch (the broken line). Taking $x_f = 1$, the bifurcation of the daughter branch from the mother branch still occurs at a distance unity from O. Taking $\nu = 2$, half the length of the mother branch (resp. the daughter branch) grows up to $(3/2)^2 = 9/4$ (resp. $2^2 = 4$). The scale ratio between the mother branch and the daughter branch then increases from $\rho = 2/(3/2) = 4/3$ to $\rho' = 4/(9/4) = 36/9 = \rho^2$.

The nature of the scale-symmetry linked to the transformations (3) follows from the fundamental property of power laws: power laws $S_{x_f,\nu}$ change scale *ratios*:

$$S_{x_f,\nu}(.) : \forall(l_1, l_2),\ \rho = \frac{l_2}{l_1} \rightarrow \rho' = \frac{S_{x_f,\nu}(l_2)}{S_{x_f,\nu}(l_1)} = \left(\frac{l_2}{l_1}\right)^\nu = \rho^\nu. \quad (4)$$

Especially, symmetry of distributions with respect to these transformations means that statistics do not make difference between scale ratios. This property stands as the analogous, for scale ratios, of the concept of scale invariance for scale: it corresponds to a *scale-ratio* invariance.

To illustrate the nature of this scale symmetry, consider the statistical properties of deterministic fractals built on a given scale ratio. An example of this is given by a Koch curve (Fig. 3a). There, the ratio $\rho = 4/3$ between the sizes of the similar branches generated at consecutive scales stands for a definite scale ratio on which the whole geometry of the object relies. In many instances, one may guess that this scale ratio ρ plays an essential role in the statistical behaviour of the system. Then, changing ρ by applying a stretching transformation $S_{x_f,\nu}$ on branches would yield still a fractal object (Fig. 3b), but with definitely different properties: there would be no scale-ratio invariance. However, it may happen that some property does not depend on the scale-ratio ρ. This would be the case for instance for a property only relying on the *number* of small scale objects generated on a branch at each new scale generation (e.g. 4 for a Koch curve). Then changing ρ for another scale-ratio ρ' would have no implication on that particular property, so that scale-ratio invariance would be valid in this instance.

2.3.2 Statistical equivalence; scale covariance

As for scale invariance, selection of distributions satisfying this symmetry may be achieved by requiring covariance with respect to the scale transformations $S_{x_f,\nu}$. This derivation is performed in Section 3.3 below. An important difference between the two kinds of covariance however stands in the existence of *two* kinds of parameters in the scale transformations $S_{x_f,\nu}$ instead of *one* in the dilations D_λ. In particular, whereas the parameter ν is relevant to the change of scale ratio (4), the parameter x_f is not. Regarding scale-ratio invariance, x_f thus stands as a *free* parameter which can be chosen as desired.

This degree of freedom means that scale-ratio invariance is satisfied provided that, given a definite ν, there *exists* one parameter x_f such that the power law $S_{x_f,\nu}$ does not modify the statistical representation. Then distribution functions are such that they may keep unchanged *despite* the change of scale ratios: they are scale-ratio *covariant*.

It is worth noticing that these covariant distribution functions are invariant by change of scale ratios only for a judicious *choice* of parameter x_f: $x_f \equiv x_f(\nu)$. This makes an explicit difference here between *covariance* and *invariance* which contrasts with the previous case of scale-invariance where the two concepts could be confused one into the other.

2.4 Previous analyses of scale symmetries for distributions

Before addressing the determinations of the covariant distributions, it is worth reporting the previous analyses that have relied on scale symmetries in this topic. In particular, it will be instructive for the sequel to highlight their differences with the present study.

The first mathematical investigation of a possible implication of scale invariance in the SDL dates back to H.Pinckham [11]. Working on the digit distribution and assuming its invariance by dilation, he derived the SDL from a functional relationship that was solved by using the properties of the unit circle mappings. This led him to the conclusion that the only scale-invariant digit distribution is the SDL, in contrast with the derivation that will be reported in Section 3.2. The precise reasons for this disagreement are pointed out in Appendix A.1.

The most comprehensive study of the link between scale symmetries and the SDL has been undergone by T. Hill who investigated the implications of both scale-invariance and base-invariance [3]. The latter symmetry is defined as an invariance by change of base of numerotation, i.e. $x = ab^n,\ n \in \mathcal{Z}, a \in [1, b[\rightarrow x' = ab'^n,\ n \in \mathcal{Z}$. The essential differences of the analyses of Hill with that reported hereafter are the following:

- The change of base corresponds to a discontinuous transformation in contrasts with the stretching transformation $S_{x_f,\nu}$. In particular, base change leaves a gap of values in between $x' = bb'^n$ and $x' = b'^{n+1}$.
- Base-invariance corresponds to a stretching transformation $S_{a,\nu}(.)$ with a fixed $\nu = \ln(b')/\ln(b)$ but a parameter a that *depends* on the point x that is considered. This implies that scale ratios $\rho = x_2/x_1$ change, but in a way that *depends* on scales: $\rho = x_2/x_1 \rightarrow \rho' = x'_2/x'_1 = \rho^\nu\ (a_2/a_1)^{1-\nu}$ where $x_1 = a_1 b^n$, $x_2 = a_2 b^m$, $\nu = \ln(b')/\ln(b)$ and where a_2/a_1 exhibits the scale dependence. As the change of scale-ratios is not uniform with respect to scales, base-invariance cannot then be considered as a scale-ratio invariance.
- The scale symmetries considered in [3] address the distribution of the most significant *digits* of data belonging to *unbounded* sets: the Borel sets $B = \cup[x_1 b^n, x_2 b^n[$ where $n \in \mathcal{Z}$ and $(x_1, x_2) \in [1, b[$. These sets actually correspond to those positive numbers that have their most significant part in base b in between x_1 and x_2. However, the fact that they extend from 0 to ∞ whereas the probability density function $\pi(x) \propto 1/x$ (41) that underlines the SDL is not normalizable on the set $[0, \infty[$ ($\int_a^b \pi dx = \pm\infty$ for $a = 0$ or $b = \infty$) questions the relevance of an unbounded data set for dealing with probabilities in this problem (see App. A.5). In particular, as the normalization factors of distributions may diverge in the limit of infinite or vanishing data (as it does for the sought SDL) unbounded data sets forbid dealing with probability density functions, for instance when considering the effect of dilations. As shown in Appendix A.1, this actually prevents selecting other scale-invariant distributions than the SDL.
- Another difference stands in the nature of the symmetries that are considered. Base-invariance is a symmetry of the representation of numbers, whereas scale-ratio covariance is a symmetry of data themselves, irrespective of their kind of representation. The latter property will enable us to easily transpose this symmetry of data into a symmetry of system.
- The last major difference stands in the concept of scale symmetry: Hill uses it as a property of *invariance* whereas we shall use it as a property of *covariance*. In particular, the fact that the concept of base-invariance involves only *one* parameter (i.e. parameter b above) makes its covariance actually reduces to an invariance. By comparison, the fact that scale-ratio invariance involves *two* parameters (x_f, ν) in the present context makes covariance and invariance two definitely different concepts.

Following these differences, it will be no surprise that opposite conclusions will be reached regarding both the nature of symmetric distributions and their link to the SDL. In particular, whereas Hill concludes that scale-invariance *solely* selects the SDL and *implies* base invariance, we shall show that scale-invariance and scale-ratio invariance are two *distinct* symmetries, separately satisfied by many *other* distributions than the SDL, but sharing a *single* common distribution: the SDL (Sect. 5).

Recently, multiplicative process [14] and stochastic systems [15] have been proposed for supporting the widespread occurrence of the SDL. In particular, simulations of both the evolution of a flat distribution submitted to a multiplicative noise [14] and of dynamical systems [15] have revealed an attraction towards the SDL. In comparison, we shall not be interested here in a particular stochastic phenomenon compatible with the SDL but in determining the deep structures of this law that all the phenomena which support it have to meet.

3 The symmetric distributions

We derive in this section the distributions satisfying either scale-invariance or scale-ratio invariance. For this, we first define below a relevant framework and then focus attention on both symmetries.

3.1 Probability on semi-closed intervals

3.1.1 Motivations

Although Newcomb's law deals with numerical sets whose bounds are not prescribed, it would be unwise to search for deriving this distribution on the unbounded set $\Re$ made by the real numbers. The reason for this is that unbounded sets may not be probabilizable, in the sense that normalized histograms may *not* converge to a *normalizable* distribution in the limit of a large number of data.

An example of this is given by the chance of picking up a given positive integer q at random within the infinite set $\mathcal{N}$ made by the positive integers. Here, the term "at random" means that, whatever it is, the chance of picking up an integer is the same for all integers. Accordingly, if this chance corresponds to a probability $p(q)$, this probability can only be a constant C: $\forall q;\ p(q) = C$. However, *no* value of C is compatible with the normalization constraint $\sum p(q) = 1$ for $q \in \mathcal{N}$: this stochastic process is not probabilizable!

This pitfall, which is responsible for many paradoxes in probability theory [19], stresses the fact that normalized histograms of integers picked up on the infinite set $\mathcal{N}$ actually do not converge in the limit of a large number of data. The reason for this is that, whatever the number q picked up, there always be *infinitely* more chance to pick-up the next numbers r *above* q ($r \in [q, \infty]$) than below ($r \in [0, q[$). However, this argument would no longer be relevant if the picked-up integers were confined to a compact sub-set of $\mathcal{N}$, e.g. $[0, N] \cap \mathcal{N}$. Then normalized histograms would actually converge to a normalizable distribution, e.g. $\forall q \in [0, N] \cap \mathcal{N}, p(q) = 1/N$.

To avoid a possible breaking of the concept of probability (see App. A.5), we therefore consider hereafter a *semi-closed* interval $[a, b[$ bounded by *unprescribed* real numbers a, b. We then introduce the probability p that a data belonging to the interval $[a, b[$ belongs also to the sub-interval $[x_i, x_j[\in [a, b[$: $p \equiv p(x_i, x_j \mid a, b)$. Our basic assumption will be that such probability exists, even if it may not be extended to the infinite limits $a \to -\infty$, $b \to \infty$ or to some definite bound (e.g. $a = 0$), as exemplified above by the probability of picking-up numbers.

3.1.2 Conditional probability

By definition, probability $p(x_i, x_j \mid a, b)$ is the probability that data x belongs to $[x_i, x_j[$, once it is granted that it belongs to $[a, b[$. It is therefore a conditional probability.

Transitivity implies:

$$\forall [x_i, x_j[\ \subset \ [a, b[\ \subset \ [c, d[$$
$$p(x_i, x_j \mid a, b) * p(a, b \mid c, d) = p(x_i, x_j \mid c, d)$$

so that

$$\forall (x_i, x_j) \ \in \ [a, b[^2 \ , \ p(x_i, x_j \mid a, b) = \frac{f(x_i, x_j)}{f(a, b)} \tag{5}$$

with $f(x_i, x_j) = p(x_i, x_j \mid c, d)$ for fixed values c, d.

Notice that changing c or d only changes the function $f(.,.)$ by a constant prefactor. As this has no implication on the probability $p(.,. \mid a, b)$, we shall omit noticing the parameters c, d in the following.

3.1.3 Additivity

Consider the intervals $[x_i, x_j[$ and $[x_j, x_k[$. As they are disjoint intervals with a union equal to the interval $[x_i, x_k[$, they yield an additive property for the corresponding probabilities:

$$\forall (x_i, x_j, x_k) \in [a, b[^3 \ ,$$
$$p(x_i, x_k \mid a, b) = p(x_i, x_j \mid a, b) + p(x_j, x_k \mid a, b)$$

This, together with relation (5), yields additivity of function $f(.,.)$:

$$\forall (x_i, x_j, x_k) \ , \ f(x_i, x_k) = f(x_i, x_j) + f(x_j, x_k) \tag{6}$$

Fixing x_k and introducing $g(x) = f(x, x_k)$, we obtain:

$$\forall (x_i, x_j) \ , \ f(x_i, x_j) = g(x_j) - g(x_i) \tag{7}$$

and, from relation (5):

$$p(x_i, x_j \mid a, b) = \frac{g(x_j) - g(x_i)}{g(b) - g(a)}. \tag{8}$$

Here too, as changing x_k only changes function $g(.)$ by a constant with no implication on the probability $p(.,. \mid a, b)$, we shall omit noticing it in the following.

Our problem now reduces to determining the functions $g(.)$ relevant to scale-invariance or to scale-ratio invariance, up to an irrelevant constant and an irrelevant prefactor.

3.2 Scale invariance

The property of statistical scale-invariance implies the covariance of normalized distributions by arbitrary scale dilation D_λ. Here, scale dilation mimics a change of scale units, so that it must apply to *all* scale measurements. These include not only the bounds x_i, x_j of the sub-interval in which data are sought, but *also* the bounds a, b of the interval in which they are assumed to lie.

Covariance then yields the following criterion for scale-invariant distribution:

$$\forall \lambda \ , \ \forall (x_i, x_j, a, b),$$
$$p(D_\lambda(x_j), D_\lambda(x_i) \mid D_\lambda(a), D_\lambda(b)) = p(x_j, x_i \mid a, b)$$

that we rewrite in the following formal condensed form:

$$\forall \lambda \ , \ p \circ D_\lambda = p. \tag{9}$$

Criterion (9) states that a scale-invariant distribution p is compatible with *any* dilation of variables. Thus, observing such distribution cannot give any information on the scale units with which the variables have been measured. From this ambiguity follows the statistical equivalence of all units regarding this distribution.

From relation (8), criterion (9) yields a constraint on function $g(.)$:

$$\forall \lambda \ , \ \forall (x_i, x_j, a, b) \ ,$$
$$\frac{g(\lambda x_j) - g(\lambda x_i)}{g(\lambda b) - g(\lambda a)} = \frac{g(x_j) - g(x_i)}{g(b) - g(a)}. \tag{10}$$

This implies that the following expressions only depend on λ via functions labelled $h(.)$ and $k(.)$:

$$\forall \lambda, \ \exists h(.), \ \exists k(.),$$
$$\forall (x_i, x_j), \ \frac{g(\lambda x_j) - g(\lambda x_i)}{g(x_j) - g(x_i)} = h(\lambda)$$
$$\forall x, \ g(\lambda x) = h(\lambda) \ g(x) + k(\lambda). \tag{11}$$

The general derivation of the solutions of criterion (11) is reported in Appendix A.6. It only assumes the existence of a distribution p, not everywhere discontinuous. However, for the sake of simplicity of the derivation, we assume below that p is differentiable with respect to x_i, x_j or, equivalently, that $g(x)$ is differentiable with respect to x. This assumption will be found not to restrict the set of solutions.

Differentiation of criterion (11) with respect to x gives:

$$\forall \lambda \neq 0, \ \forall x, \ g'(\lambda x) = l(\lambda) \ g'(x) \tag{12}$$

where $l(\lambda) = h(\lambda)/\lambda$ and where primes denote differentiation with respect to x at fixed λ.

Differentiation of relation (12) with respect to λ gives:

$$\forall\lambda \neq 0,\ \forall x,\ x\, g''(\lambda x) = \dot{l}(\lambda)\, g'(x) \tag{13}$$

where dots denote differentiation with respect to λ at fixed x.

Combining relations (12) and (13) gives the following constant expressions:

$$\forall\lambda \neq 0,\ \forall y = \lambda x,\ \frac{y\, g''(y)}{g'(y)} = \frac{\lambda\, \dot{l}(\lambda)}{l(\lambda)} = e - 1 \tag{14}$$

where e is a constant.

Integrating (14) yields two cases:

– $e \neq 0$

$$g(x) = \alpha\, x^e + \beta \tag{15}$$

– $e = 0$

$$g(x) = \alpha\, \ln(|x|) + \beta \tag{16}$$

where (α, β) are constants. They correspond to the following scale-invariant distributions:

– $e \neq 0$

$$p_e(x_i, x_j \mid a, b) = \frac{x_j^e - x_i^e}{b^e - a^e} \tag{17}$$

– $e = 0$

$$p_0(x_i, x_j \mid a, b) = \frac{\ln(|x_j|) - \ln(|x_i|)}{\ln(|b|) - \ln(|a|)}. \tag{18}$$

Notice that, beyond the formal split at $e = 0$, the space of solutions is continuous in the sense that p_0 is simply the limit of p_e for $e \to 0$. In addition, the relevance of the restriction to a bounded range of data $[a, b[$ is emphasized by the divergence or the vanishing of distributions (17, 18) in the limit of zero or infinite a, b, x_i or x_j (see App. A.5).

3.3 Scale-ratio invariance

The property of scale-ratio invariance implies the covariance of normalized distributions by *arbitrary* change of scale-ratios:

$$\rho \to \rho^\nu. \tag{19}$$

It is worth stressing here that scale-ratio changes (19) are only parametrized by ν so that a given scale-ratio transformation (4) is driven by *any* power law transformations $S_{x_f,\nu}(.)$ involving the desired value of ν:

$$S_{x_f,\nu}(.):\ x \to x_f \left(\frac{x}{x_f}\right)^\nu. \tag{20}$$

Therefore, the requirement of scale-ratio covariance asks that, for *any* ν, the invariance by at least *one* power law change of variables $S_{x_f,\nu}(.)$ is satisfied for at least *one* x_f. Here too, variable change must apply to *all* scale measurements, i.e. to x_i, x_j, a and b.

We thus obtain the following formal criterion for scale-ratio invariant distribution:

$$\begin{aligned}&\forall\nu, \exists x_f;\ \forall(x_i, x_j, a, b)\ ,\\ &p(S_{x_f,\nu}(x_j), S_{x_f,\nu}(x_i)\ \mid\ S_{x_f,\nu}(a), S_{x_f,\nu}(b))\\ &\qquad = p(x_j, x_i \mid a, b)\end{aligned}$$

that we rewrite in the following formal condensed form:

$$\forall\nu\ , \exists x_f;\ p \circ S_{x_f,\nu} = p. \tag{21}$$

Criterion (21) means that a scale-ratio invariant distribution p is compatible with *any* change of scale ratio $\rho \to \rho^\nu$. Thus, observing such distribution cannot give *any* information on the status of scale-ratio measurements and, therefore, *any* relevance to their values. From this ambiguity follows the statistical equivalence of scale ratios regarding this kind of distributions.

Changing variable x for its logarithm $X = \ln(|x|)$, we notice that transformation $S_{x_f,\nu}(.)$ corresponds to an affine transformation $A_{M,\nu}$ of variable X:

$$S_{x_f,\nu}(.)\ :\ X = \ln(|x|) \to \nu X + M = A_{M,\nu}(X) \tag{22}$$

with:

$$M = (1 - \nu)\ln(|x_f|) \tag{23}$$

with x_f possibly dependent on ν at this stage.

Let us label $P(.,. \mid .,.)$ and $G(.)$ the expression of functions $p(.,. \mid .,.)$ and $g(.)$ in logarithmic variables $X = \ln(|x|)$:

$$P \circ \ln = p;\ \ G \circ \ln = g. \tag{24}$$

The equivalent covariant criterion for P writes:

$$\forall\nu\ , \exists M;\ P \circ A_{M,\nu} = P. \tag{25}$$

In a way similar as in Section 3.2, this yields from relation (8) a constraint on function $G(.)$:

$$\begin{aligned}&\forall\nu,\ \exists M(.); \exists H(.),\ \exists K(.)\ ;\\ &\qquad G(\nu X + M(\nu)) = H(\nu)\ G(X) + K(\nu).\end{aligned} \tag{26}$$

The only difference with criterion (11) is the additional degree of freedom $M(.)$. General derivation for well defined distribution p, not everywhere discontinuous, is reported to Appendix A.7. However, for the sake of simplicity, we assume below that P is differentiable with respect to both $X_i = \ln(|x_i|)$, $X_j = \ln(|x_j|)$ or, equivalently, that $G(X)$ is differentiable with respect to X. This will be found not to restrict the set of solutions.

Differentiation of criterion (26) with respect to X gives:

$$\forall\nu \neq 0,\ \forall X,\ G'(\nu X + M) = L(\nu)\ G'(X) \tag{27}$$

where $L(\nu) = H(\nu)/\nu$ and where primes denote differentiation with respect to X at fixed ν.

Differentiation of relation (27) with respect to ν gives:

$$\forall\nu \neq 0,\ \forall X,\ (X + \dot{M})\ G''(\nu X + M) = \dot{L}(\nu)\ G'(X) \tag{28}$$

where dots denote differentiation with respect to ν at fixed X.

Combining relations (27) and (28), we obtain:

$$\forall \nu \neq 0,\ \forall Y = \nu X + M,\ \frac{G''(Y)}{G'(Y)} = \frac{R}{Y - N} \tag{29}$$

where $R(\nu) = \nu \dot{L}(\nu)/L(\nu)$ and $N(\nu) = M(\nu) - \nu \dot{M}(\nu)$.

The fact that function $P(.,. \mid .,.)$, and thus function $G(.)$, does not explicitly depend on parameter ν implies that both R and N are *constant*. In particular, the latter property means that $M(\nu)$ is affine: $M(\nu) = m + n\nu$ where m and n are constants. Relation (23) then implies $M(1) = 0$, $n = -m$ and $\ln(|x_f|) = m$, so that x_f is *independent* of ν. This means that *all* the fixed points of the power law transformations ensuring covariance are the *same*. From now on, we shall label x_c their common value: $\forall \nu, x_f(\nu) = x_c$.

Integration of (29) at *fixed* ν provides two cases which, back to variable x and function $g(.)$, read:

– $f \neq 0$

$$g(x) = \alpha\ (\ln(|x/x_f|)^f + \beta \tag{30}$$

– $f = 0$

$$g(x) = \alpha\ \ln(|\ln(|x/x_f|)|) + \beta. \tag{31}$$

Here f, α, β, x_f are independent of Y as a result of the integration and of the definition (3). They are also independent of ν since neither function $g(.)$ nor function $G(.)$ depend on this parameter. They are therefore actual constants.

Relations (30, 31) correspond to the following family of scale-ratio invariant distributions:

– $f \neq 0$

$$p_{f,x_c}(x_i, x_j \mid a, b) = \frac{[\ln(|x_j/x_c|)]^f - [\ln(|x_i/x_c|)]^f}{[\ln(|b/x_c|)]^f - [\ln(|a/x_c|)]^f} \tag{32}$$

– $f = 0$

$$p_{0,x_c}(x_i, x_j \mid a, b) = \frac{\ln[|\ln(|x_j/x_c|)|] - \ln[|\ln(|x_i/x_c|)|]}{\ln[|\ln(|b/x_c|)|] - \ln[|\ln(|a/x_c|)|]}. \tag{33}$$

Notice that, beyond the formal split at $e = 0$, the space of solutions is continuous in the sense that p_{0,x_c} is simply the limit of p_{f,x_c} for $f \to 0$. Notice also that no distribution of the kind of (32, 33) could exist in a set extending to infinity, i.e. $b = \infty$, or including 0, e.g. $a = 0$ (see App. A.5). Here too, this stresses the relevance of the restriction to a bounded range of data $[a, b[$.

The value x_c stands as a characteristic scale of the distributions (32, 33). The fact that it also corresponds to the common value of the fixed points $x_f(\nu)$ of the covariant transformations $S_{x_f,\nu}(.)$ means that distributions (32, 33) are covariant by the stretching transformations:

$$S_{x_c,\nu}(x) = x_c (x/x_c)^\nu \tag{34}$$

as can be checked straightforwardly.

4 The significant digit law

In Section 2, the analysis of the context in which the SDL appears led us to conjecture that it could refer to the most symmetric distributions regarding scales. The identification in Section 3 of the distributions that are scale-invariant or scale-ratio invariant now enables us to check this conjecture directly.

We show below that the SDL actually corresponds to the distribution common to the set of scale-invariant distributions and to the set of scale-ratio invariant distributions. We then investigate the attractor properties of this specific distribution by dilation, stretch and merge.

4.1 The SDL: nature and symmetries

4.1.1 Distribution of the most significant digit

The distribution of the most significant digit of data expressed in a given base simply follows from the conditional probability of data occurrence considered in previous sections. For instance, for a data set ranging from 10^{-N} to 10^N and expressed in the decimal base, the distribution $P(s \mid 10^{-N}, 10^N)$ of the most significant digit s, $1 \leq s \leq 9$, is obtained by summing the different probabilities of finding a data in the ranges $[s10^n, (s+1)10^n[$, $1 \leq n \leq N-1$:

$$P(s \mid 10^{-N}, 10^N) = \sum_{n=-N}^{N-1} p\ (s10^n, (s+1)10^n \mid 10^n, 10^{n+1}) * p\ (10^n, 10^{n+1} \mid 10^{-N}, 10^N). \tag{35}$$

Notice that the relationship (35) could be easily generalized to any other base b by replacing the argument 10 by b.

4.1.2 Continuous SDL

The significant digit law (1) expressed by Newcomb actually addressed the distribution of the first significant digit in the decimal base. It has been extended to the first two significant digits by Newcomb [1] and then to the n first digits $(d_1, d_2, \ldots, d_n) \in \{1, \ldots, 9\}^n$ by Hill [3]. This yielded the generalized n-digit SDL:

$$P_n(\sigma_n) = \frac{\ln(\sigma_n + 10^{-n}) - \ln(\sigma_n)}{\ln(10) - \ln(1)} \tag{36}$$

where $\sigma_n = [d_1 + 10^{-1}d_2 + \ldots + 10^{-n}d_n]$ and where $P_n(\sigma_n)$ denotes the probability of occurrence of the most significant part σ_n of data mantissae.

The n-digit SDL (36) is a *discrete* distribution law. However, in the limit of infinite n, it yields a *continuous* distribution that may be named *continuous* significant digit law (CSDL):

$$\forall (x, y) \in [1, 10[^2\ ,\ P(x, y) = \frac{\ln(y) - \ln(x)}{\ln(10) - \ln(1)} \tag{37}$$

where $P(x,y)$ expresses the probability of occurrence of mantissa σ in between x and y, $1 \leq x < 10$, $1 \leq y < 10$. Of course, the CSDL contains the SDL as a restriction to $x = s$, $y = s+1$ and $s \in \{1, \dots, 9\}$.

The stability of the SDL with respect to data stretching [13] as well as direct investigation [3] shows that the base of the numerotation b is of no importance with regards to the SDL. This yields us to generalize the CSDL as:

$$\forall (x,y) \in [1,b[^2 \ , \ \forall b > 0 \ , \ P(x,y,b) = \frac{\ln(y) - \ln(x)}{\ln(b) - \ln(1)}. \tag{38}$$

Notice finally hat, for data belonging to a given range $[b^{-N_1}, b^{N_2}[$, the connection between the CSDL and conditional probabilities simply follows from the relationship:

$$\begin{aligned} &\forall (x,y) \in [1,b[^2, \ \forall b > 0 \\ &P(x,y,b \mid b^{-N_1}, b^{N_2}) = \sum_{n=-N_1}^{N_2-1} p(xb^n, yb^n \mid b^n, b^{n+1}) \\ &\qquad\qquad * \ p(b^n, b^{n+1} \mid b^{-N_1}, b^{N_2}). \end{aligned} \tag{39}$$

4.1.3 Equivalence between Continuous SDL and logarithmic distribution

Assuming that the CSDL is valid in any range, we can apply (38) in the two ranges $[b^{-N_1}, b^{N_2}[$ and $[b^{-N_1}, b^{N_2-1}[$. We obtain the following independence with respect to data ranges:

$$\begin{aligned} &\forall (x,y) \in [1,b[^2 \ , \ \forall b > 0 \\ &P(x,y,b \mid b^{-N_1}, b^{N_2}) = P(x,y,b \mid b^{-N_1}, b^{N_2-1}) \\ &\qquad = P(x,y,b). \end{aligned}$$

On the other hand, separating in relation (39) the term $n = N_2 - 1$ of the summation from the others, we obtain:

$$\begin{aligned} &\forall (x,y) \in [1,10[^2 \ , \ \forall b > 0, \forall (N_1, N_2) \\ &P(x,y,b \mid b^{-N_1}, b^{N_2}) = \\ &P(x,y,b \mid b^{-N_1}, b^{N_2-1}) * p(b^{-N_1}, b^{N_2-1} \mid b^{-N_1}, b^{N_2}) \\ &+ p(xb^{N_2-1}, yb^{N_2-1} \mid b^{N_2-1}, b^{N_2}) * p(b^{N_2-1}, b^{N_2} \mid b^{-N_1}, b^{N_2}) \end{aligned}$$

and finally:

$$\begin{aligned} &\forall (x,y) \in [1,10[^2 \ , \ \forall b > 0, \forall (N_1, N_2) \\ &p(xb^{N_2-1}, yb^{N_2-1} \mid b^{N_2-1}, b^{N_2}) = P(x,y,b) \\ &\qquad = \frac{\ln(y) - \ln(x)}{\ln(b) - \ln(1)}. \end{aligned}$$

As b and N_2 are arbitrary, this corresponds to the logarithmic distribution (18) with $x_i \equiv xb^{N_2-1}$, $x_j \equiv yb^{N_2-1}$, $a \equiv b^{N_2-1}$ and $b \equiv b^{N_2}$.

Conversely, assume that the logarithmic distribution is valid in any ranges. Then relation (39) straightforwardly yields the CSDL (38). The continuous significant digit law (38) is thus *equivalent* to the logarithmic distribution (18).

4.1.4 Scale invariance and scale-ratio invariance: logarithmic distribution and SDL

Comparing relations (17, 18, 32, 33) shows that there exists a *single* distribution satisfying *both* scale-invariance and scale-ratio invariance: the logarithmic distribution:

$$\begin{aligned} p_0(x_i, x_j \mid a, b) &= p_{1,x_c}(x_i, x_j \mid a, b) \\ &= \frac{\ln(|x_j|) - \ln(|x_i|)}{\ln(|b|) - \ln(|a|)}. \end{aligned}$$

This, together with the above results, shows that there is equivalence between the couple of symmetries of scale-invariance *plus* scale-ratio invariance, the logarithmic distribution and the SDL. In particular, in contrast with some analyses [3,11], scale-invariance only is not sufficient to select the SDL: scale-ratio invariance is *required* too.

Notice that the present selection of the SDL relies on symmetries of *data* distribution. A selection of the SDL from the symmetries of *digit* distribution is reported in Appendix A.1. It also shows that both the scale invariance and the scale-ratio invariance are needed to recover the SDL.

4.1.5 Probability density functions

Although the data sets to which the present study is dedicated are finite, it will be useful to interpret the form of conditional probabilities $p(x_i, x_j \mid a, b)$ in terms of probability density functions. Assuming a continuous data distribution, we thus introduce the probability density function (p.d.f.) $\pi(x \mid a, b)$ derived from the conditional probability $p(x_i, x_j \mid a, b)$ as:

$$\pi(x \mid a, b) = \frac{dp}{d\epsilon}(x, x+\epsilon \mid a, b) \tag{40}$$

In particular, the p.d.f. of the SDL, noticed π_s, reads:

$$\pi_s(x \mid a, b) = \frac{1}{\ln(b) - \ln(a)} \frac{1}{x}. \tag{41}$$

It thus corresponds the following property:

$$\frac{d}{dx}[x\pi(x \mid a, b)] = 0. \tag{42}$$

This provides a simple criterion for identifying it.

4.2 Attraction towards the SDL

Being symmetric, the SDL is a fixed point of the scale transformations changing the scales (i.e. the dilations (2)) or the scale-ratios (i.e. the power law transformations (3)). In this context, a relevant question emerges as to whether iterating these transformations on other distributions makes them approach or escape the vicinity of the only invariant distribution, the SDL. This amounts to determining whether the SDL is an attractor or a repulsor in the distribution space for dilations or for stretching transformations.

Two different points of view may be taken for considering the new distribution resulting from dilation or stretch. Their difference refers as to whether one analyses the *whole* new data set or only its intersection with a *fixed* bounded observation window. This turns out considering whether data sampling occurs *before* or *after* the iterations of dilation or stretch. The first case corresponds to an observer that changes iteratively the data he has sampled once: he then works on a *fixed* data set that is latter changed. The second case corresponds to an observer which samples in a given range a phenomenon that is iteratively distorted by dilations or stretches: he then works with the *varying* parts of the initial data set that fits within his observation window.

We shall refer to these different cases as to "fixed data set" and "fixed observation window". Their main difference refers to the origin of the attraction: the intrinsic mixing property of transformations in the first case (Sect. 4.2.2); the form of the p.d.f. of the initial data set in the latter case (Sect. 4.2.1).

4.2.1 Fixed observation window

Data are changed by dilations or stretch but observed in a *fixed* bounded observation window $[a, b[$, $0 \leq a,\ b < \infty$.

Dilation

Consider a distribution $p(x_i, x_j \mid a, b)$ and apply a dilation of data: $x \to D_\lambda(x) = \lambda x$. One obtains a dilated distribution $D_\lambda[p](x_i, x_j \mid a, b)$ which is related to the initial one by:

$$D_\lambda[p](D_\lambda(x_i), D_\lambda(x_j) \mid D_\lambda(a), D_\lambda(b)) = p(x_i, x_j \mid a, b) \tag{43}$$

or:

$$D_\lambda[p](x_i, x_j \mid a, b) = p(D_\lambda^{-1}(x_i), D_\lambda^{-1}(x_j) \mid D_\lambda^{-1}(a), D_\lambda^{-1}(b)). \tag{44}$$

Iterating this, one obtains:

$$\begin{aligned} D_\lambda[p]^n(x_i, x_j \mid a, b) &= p(D_\lambda^{-n}(x_i), \\ &\qquad D_\lambda^{-n}(x_j) \mid D_\lambda^{-n}(a), D_\lambda^{-n}(b)) \\ &= p(x_i\lambda^{-n}, x_j\lambda^{-n} \mid a\lambda^{-n}, b\lambda^{-n}). \end{aligned}$$

In particular, as function $D_\lambda(.)$ is dilating, i.e. $\lambda > 1$, iterating it backwards makes the data that belong to the bounded range $[a, b[$ uniformly converge to zero: $D_\lambda^{-n} \to 0$ for $n \to \infty$. This implies that the distribution function $D_\lambda[p]^n$ is determined by the asymptotic expression of the pdf $\pi(. \mid ., .)$ of the probability p in the vicinity of 0.

In particular, if $\pi(x \mid ., .) \approx x^{-1}$ in the vicinity of 0, the dilated distribution $D_\lambda[p]^n$ converges towards the SDL for $n \to \infty$:

$$\begin{aligned} D_\lambda[p]^n(x_i, x_j \mid a, b) &\to_{n\to\infty} \int_{x_i\lambda^{-n}}^{x_j\lambda^{-n}} \pi(x \mid a\lambda^{-n}, b\lambda^{-n})dx \\ &\to_{n\to\infty} \frac{\ln(x_j) - \ln(x_i)}{\ln(b) - \ln(a)}. \end{aligned}$$

More generally, the way $\pi(x \mid a, b)$ tends to 0 with x, a and b, determines the limit expression of $D_\lambda[p]^n$. In particular, if $\pi(x \mid a, b) \approx x^{e-1}$ in the vicinity of 0, the dilated distribution $D_\lambda[p]^n$ converges to:

$$D_\lambda[p]^n(x_i, x_j \mid a, b) \to_{n\to\infty} \frac{x_j^e - x_i^e}{b^e - a^e} \tag{45}$$

i.e. to a *scale-invariant* distribution that is *not* invariant by scale ratio changes.

Dilating data thus turns out performing a zoom of the distribution around 0, so that the infinitely dilated distribution behaves as the starting distribution does around 0. Accordingly, the basin of attraction of distributions for iterated dilations is determined by the shape of these distributions for vanishingly small data.

Stretch

Consider a distribution $p(x_i, x_j \mid a, b)$ and apply a power law stretch of data (3). One then obtains a stretched distribution $S_{x_f,\nu}[p](x_i, x_j \mid a, b)$ which is related to the initial one by:

$$\begin{aligned} S_{x_f,\nu}[p](S_{x_f,\nu}(x_i), S_{x_f,\nu}(x_j) \ \mid\ S_{x_f,\nu}(a), S_{x_f,\nu}(b)) \\ = p(x_i, x_j \mid a, b). \end{aligned}$$

Iterating this, one obtains:

$$\begin{aligned} &S_{x_f,\nu}^n[p](x_i, x_j \ \mid\ a, b) \\ &\qquad = p[S_{x_f,\nu}^{-n}(x_i), S_{x_f,\nu}^{-n}(x_j) \mid S_{x_f,\nu}^{-n}(a), S_{x_f,\nu}^{-n}(b)]. \end{aligned}$$

However, as function $S_{x_f,\nu}(.)$ is stretching, $\nu > 1$, $m = 1/\nu < 1$, iterating it backwards makes data converge exponentially slowly towards its fixed point x_f:

$$\begin{aligned} S_{x_f,\nu}^{-n}(x) &= x_f\ |x/x_f|^{m^n} \\ &= x_f\ \exp[m^n\ \ln(|x/x_f|)] \\ &\to x_f \quad \text{for} \quad n \to \infty. \end{aligned} \tag{46}$$

This implies that the distribution function $S_{x_f,\nu}^n[p]$ is determined by the asymptotic expression of the p.d.f. $\pi(. \mid ., .)$ of the probability p in the vicinity of x_f.

In particular, for $\pi(x \mid a, b) \approx \alpha(x - x_f)^{f-1}$ in the vicinity of x_f and the asymptotic trend (46), we obtain, for $n \to \infty$:

$$\begin{aligned} &S_{x_f,\nu}^n[p](x_i, x_j \ \mid a, b) \to_{n\to\infty} \\ &\qquad \times \frac{[S_{x_f,\nu}^{-n}(x_j) - x_f]^f - [S_{x_f,\nu}^{-n}(x_i) - x_f]^f}{[S_{x_f,\nu}^{-n}(b) - x_f]^f - [S_{x_f,\nu}^{-n}(a) - x_f]^f} \\ &\qquad \to_{n\to\infty} \frac{[\ln(|x_j/x_f|)]^f - [\ln(|x_i/x_f|)]^f}{[\ln(|b/x_f|)]^f - [\ln(|a/x_f|)]^f}. \end{aligned}$$

The iterated stretched distribution thus converges towards a SRI distribution (32). For $f \neq 1$, this distribution is not invariant by scale dilation; for $f = 1$, it is, since it corresponds to the SDL.

The infinitely stretched distribution thus behaves as the starting distribution does around the fixed point x_f of the transformation. Accordingly, the basin of attraction of distributions for iterated stretching is determined by the shape of these distributions around x_f.

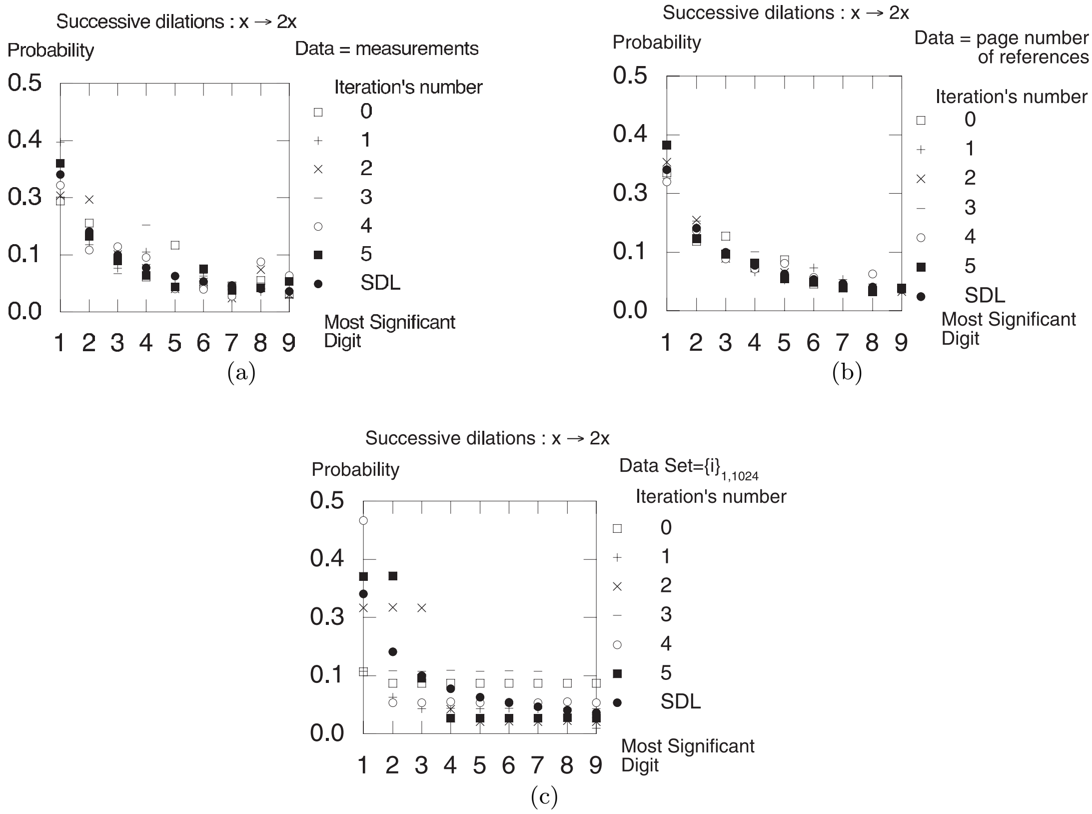

Fig. 4. Sketch of the effects of iterated dilations on distributions. Data correspond to: (a) the measurements reported in the Vol. 22 of Eur. Phys. J. B [4]. (b) the index of the first page of the cited articles in this volume. (c) the integer series: $1 \leq i \leq 1024$.

4.2.2 Fixed data set

Here, data set is changed by dilations or stretch and then analysed as a *whole*.

We draw attention below on the distribution of the mantissa σ of data d in base b: $\sigma = db^{-n}$, $\sigma \in [a, b[$. As the SDL corresponds to an equiprobability of the logarithms of mantissae, i.e. to a uniform distribution of $\log(\sigma) = \log d \equiv \log b$, it may be viewed as the expression of the equiprobability of the log-data when mapped on the unit circle: $d \rightarrow \theta = 2\pi \log(d)/\log b$.

Following this property, we find it convenient to consider the distribution of log-data, $\theta = 2\pi \log(d)/\log b$, on the unit circle to better address the effect of dilations or stretch on data statistics.

Dilation

A dilation D_λ of data corresponds to a uniform translation $\log d \rightarrow \log d + \log \lambda$ of log-data. It has then *no* implication on the level of uniformity of the distribution of log-data on the unit circle. Dilation thus does not change the overall proximity of the distribution of a data set to the SDL.

Stretch

A stretch $S_{x_f,\nu}$ of data corresponds to an affine transformation $\log d \rightarrow \nu \log d + \log x_f$ of log-data. Whereas translation has no implication on the level of uniformity of the distribution, dilation yields mixing when iterated on the unit circle. This implies that iterative stretching of data yields convergence towards a uniform distribution on the log-circle, i.e. to a data set whose distribution ever more approaches the SDL.

4.2.3 Distance to the SDL

The effects of iterated dilations and stretches are shown in Figures 4–6 on three data sets: the measurements reported in the volume 22 of Eur. Phys. J. B [4], the numbers labelling the first page of the cited articles in this volume, the integer series $1 \leq i \leq 1024$. In particular, Figure 4 (resp. 5) shows the evolution of these distributions by iterated dilation (resp.stretch). However, to better evaluate the proximity to the SDL, it is useful to introduce the following distance between distributions:

$$d^2(p_i, p_j) = \sum_{s=1}^{s=9} [p_i(s) - p_j(s)]^2. \qquad (47)$$

Figure 6 then reports the evolution of the distance of the above distributions to the SDL by dilation and stretch.

Attraction to the SDL by iterative stretch is clearly evidenced on the data sets made by the index of the first page of cited papers (Fig. 6b) and by the integer series

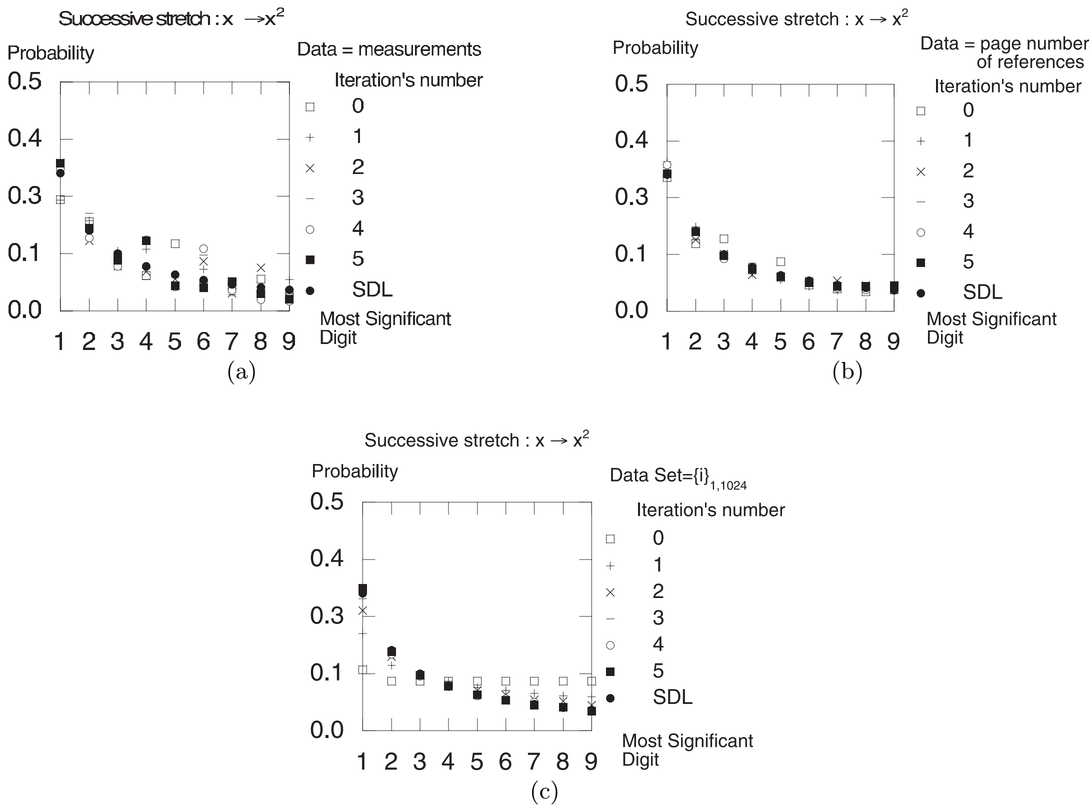

Fig. 5. Sketch of the effects of iterated stretches on distributions. Data correspond to: (a) the measurements reported in the Vol. 22 of Eur. Phys. J. B [4]. (b) the index of the first page of the cited articles in this volume. (c) the integer series: $1 \leq i \leq 1024$.

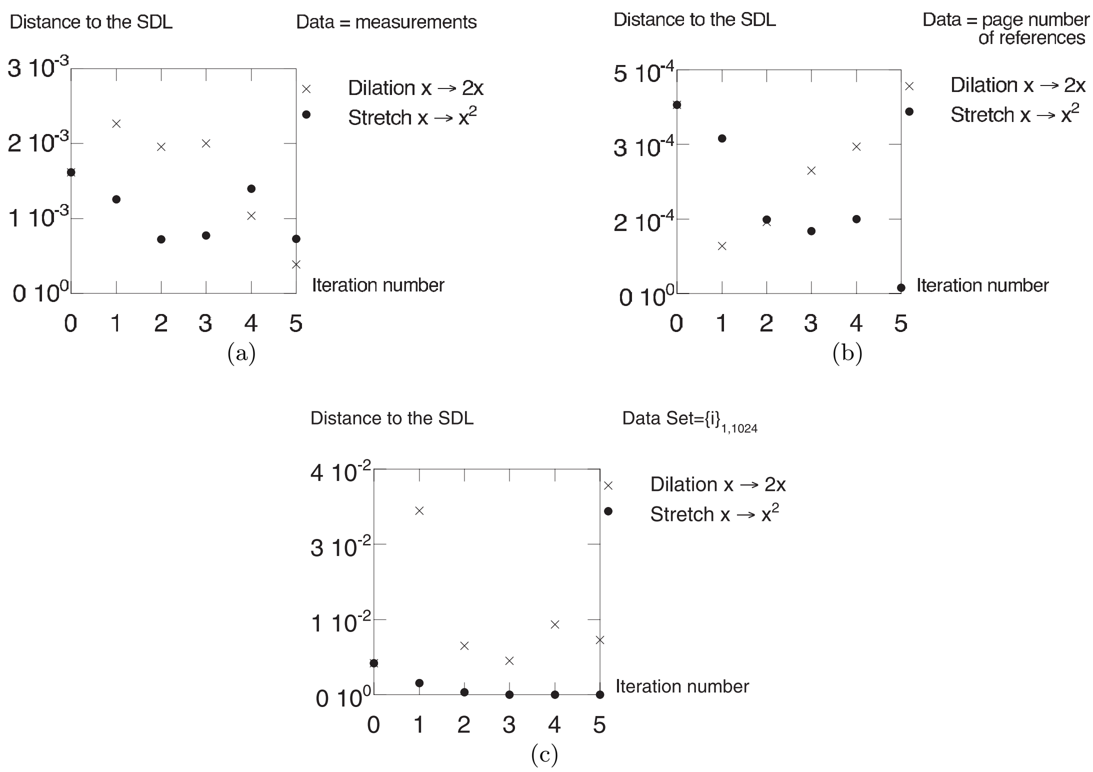

Fig. 6. Distance of distributions to the SDL distribution as defined in (47). Distributions are changed by iterative dilations or stretch. Data correspond to: (a) the measurements reported in the Vol. 22 of Eur. Phys. J. B [4]. (b) the index of the first page of the cited articles in this volume. (c) the integer series: $1 \leq i \leq 1024$.

from 1 to 1024 (Fig. 6c). It si more hardly seen on data referring to measurements (Fig. 6a), especially because of an anomalously large deviation at the fourth iteration. Its origin may be traced back to the large variations of the initial distribution from digit to digit (Fig. 4a), and especially to the large fluctuation displayed on the fifth digit. That fluctuation propagates from digit to digit on successive distributions, despite the smoothing effect of stretching. However, altogether these evolutions corroborate the attraction of distributions by stretching when the successive digit distributions of the same initial data set are considered.

By comparison, no clear tendency emerge for the evolution of distributions with iterated dilations: the distance to the SDL slightly decreases when data correspond to measurements (Fig. 6a), increases after an initial fall when data are made by the index of the first pages of the references (Fig. 6b), and slightly increases except for a large initial jump (Fig. 6c) on the data set made by positive integers up to 1024. Altogether, these behaviours corroborate the absence of attraction of dilation when the successive digit distributions of the same initial data set are considered.

4.2.4 Convergence by merge

Consider a set of distributions represented by the set of p.d.f. π_i, $i = 1, .., n$ and the distribution that results from their merging. We seek to determine the conditions for which the merged distribution may be closer to the SDL than any of its components.

We first notice that merging can actually give a distribution farther from the SDL than some of its components. In particular, merging the SDL with a non-SDL distribution of course cannot yield the SDL. Some conditions have thus actually to be achieved to make merging a convergent process towards the SDL.

To evaluate the convergence or the divergence to the SDL, we use the criterion (42) of identification of the p.d.f. of the SDL: $d/dx(x\pi_s) = 0$. It yields us to introduce a relevant distance $D(\pi_i, \pi_j)$ between distributions:

$$D(\pi_i,\pi_j)^2 = \int_a^b \left[\frac{d(x\pi_i)}{dx} - \frac{d(x\pi_j)}{dx}\right]^2 dx. \qquad (48)$$

This distance follows from a scalar product $P(\pi_i, \pi_j)$ between distributions:

$$P(\pi_i,\pi_j) = \int_a^b \frac{d(x\pi_i)}{dx}\,\frac{d(x\pi_j)}{dx} dx \qquad (49)$$

following which:

$$\begin{aligned} D(\pi_i,\pi_j)^2 &= P(\pi_i-\pi_j,\pi_i-\pi_j)\\ &= D(\pi_i,\pi_s)^2 + D(\pi_j,\pi_s)^2 - 2P(\pi_i,\pi_j). \end{aligned} \qquad (50)$$

This scalar product measures the correlation between distributions regarding characteristic scales. In particular, the domains where $d(x\pi_k)/dx = 0$, $k = i, j$, involve no characteristic scale, as the SDL; those where $d(x\pi_k)/dx \neq 0$, $k = i, j$, include characteristic scales that can interact positively or negatively depending on whether these derivatives have the same sign or a different sign. The product P then indicates whether, in average, merging the two distributions π_i, π_j, makes them reinforce ($P > 0$) or weaken ($P < 0$) their characteristic scales.

In Section 2.1, we made the guess that merging two distributions yields closer to the SDL if this comes out reducing the importance of their own features. In agreement with this guess, we shall consider distributions having negative or, at most, vanishing mean correlation: $P \leq 0$. Our problem now reduces to determining the additional criterion required to make the distance to the SDL decrease by merging.

To address this issue, we restrict ourselves to a pair of distributions, π_1, π_2. Merging them yields a distribution with a p.d.f π satisfying:

$$\pi = \gamma\pi_1 + (1-\gamma)\pi_2 \qquad (51)$$

where π_1, π_2 denote the p.d.f. of the initial distributions and γ, $0 < \gamma < 1$, a constant that depends on the number of data of each distribution. Merging yields a distribution closer to that of the SDL, π_s, if:

$$D(\pi,\pi_s) < Inf[D(\pi_1,\pi_s), D(\pi_2,\pi_s)]. \qquad (52)$$

Let us consider for simplicity that the second distribution is closer to the SDL than the first: $D(\pi_2, \pi_s) < D(\pi_1, \pi_s)$. As $d(x\pi_s)/dx = 0$, the criterion (52) simply reads:

$$\int_a^b \left[\frac{d(x\pi)}{dx}\right]^2 - \left[\frac{d(x\pi_2)}{dx}\right]^2 dx < 0 \qquad (53)$$

or, equivalently:

$$\gamma \int_a^b \left[\frac{d(x\pi_1)}{dx}\right]^2 + (\gamma-2)\int_a^b \left[\frac{d(x\pi_2)}{dx}\right]^2 + 2(1-\gamma)\int_a^b \frac{d(x\pi_2)}{dx}\,\frac{d(x\pi_1)}{dx} < 0. \qquad (54)$$

For $P \leq 0$, criterion (54) is satisfied if:

$$\gamma < \frac{2D^2(\pi_2,\pi_s)}{D^2(\pi_1,\pi_s) + D^2(\pi_2,\pi_s)}. \qquad (55)$$

For an equal merging, i.e. $\gamma = 1/2$, it is in particular valid provided that $D^2(\pi_1, \pi_s) < 3D^2(\pi_2, \pi_s)$. This loose condition shows that merging can actually decrease the distance to the SDL in many instances. In addition, criterion (54) shows that convergence may also be achieved even for weakly correlated distributions, i.e. for small negative P.

When criterion (52) is satisfied on an iterated merging of distributions, the distance of the resulting distribution to the SDL is smaller than the distance of any of the distributions that compose it. This means that merging favors convergence towards the SDL.

4.2.5 Convergence by projection in the digit space

Given a conditional probability p of occurrence of data, the corresponding digit distribution (35) may be written:

$$P(s \mid 10^{-N}, 10^{N}) = \sum_{n=1}^{N-1} D_{10^{-1}}[p]^{n}(s, s+1 \mid 1, 10) \\ p(10^{n}, 10^{n+1} \mid 10^{-N}, 10^{N}). \qquad (56)$$

Relation (56) shows that the digit law results from a merge of dilated sub-parts of the initial data distribution. In the case where the scale ratio 10 has no relevant meaning for the data set, the different sub-parts, once dilated to all lay in the range $[1, 10[$, may be uncorrelated enough to satisfy the convergence by merge to the SDL (criterion (52)). Then, the significant digit distribution will be closer to the logarithmic distribution for the whole data than it is in average in the different decades.

Notice that this property does not preclude the effect of dilation or stretch on a data distribution. In particular, whereas dilation makes distribution converges towards the SDL in any fixed decade (Sect. 4.2.1), it has no implication on the convergence towards the SDL on the whole data set (Sect. 4.2.2). The reason for this is that, whereas convergence towards the SDL is in order in the initial decades covered by the data set, new decades that increase the overall distance to the SDL arise due to data dilation. The conclusion of Section 4.2.2 is then that both effects tend to balance one the other, up to finite size and aliasing effects.

5 Scale-invariance versus scale-ratio invariance

The preceding quest for the significant digit law by symmetry arguments has revealed the definite roles of two scale symmetries: one famous, the scale-invariance, and another somewhat similar but far more ignored, the scale-ratio invariance. Both of them were actually required to select the SDL by covariance, and thus to identify the specific symmetries that characterize this special distribution.

Beyond the present topic, one may guess that these scale symmetries — including the scale-ratio invariance — could play an important role in our understanding of a number of other issues in which the concepts of scale and scale ratios are relevant [20,21]. In this spirit, the present study may be viewed as a canonical framework, simple enough to identify the implications of statistical symmetries regarding scales or scale-ratios, and rich enough to exemplify the analogies and the differences between these two concepts. Below, we develop that second aspect by addressing the concepts of characteristic scales or scale-ratios and the structure of the sets of symmetric distributions.

5.1 Characteristic scale

A characteristic scale is a scale that is singled out by the phenomenon under study, so that it may be unambiguously identified from it.

5.1.1 SI distribution: *no* characteristic scale

The property underlying the concept of scale-invariance is the absence of characteristic scale. Following it, any scale may equivalently play the role of any other in the representation of the system. This provides an operational implication of scale-invariance which may be used to identify it. This may be exemplified in relations (17, 18) by introducing an explicit normalisation scale X_s playing the role of the standard used to get data. This turns out interpreting data, e.g. x_i, as the ratio of a given quantity X_i by the standard X_s used to measure it: $x_i = X_i/X_s$. Here X_s *a priori* stands as a characteristic scale to which the resulting distribution will depend. However, for the expressions (17, 18) of scale-invariant distributions, it surprisingly happens that X_s eventually disappears from them. This confirms that its definite value is irrelevant here: SI distributions rely on no characteristic scale.

5.1.2 SRI distributions: a *single* characteristic scale

In SRI distributions (32, 33), the preceding role of the normalization scale X_s is undergone by the explicit scale x_c. Except for the specific case of the SDL ($f = 1$), the definite value of x_c then corresponds to an actual parameter of the distributions. In particular, the fact that a given SRI distribution enables to determine x_c unambiguously shows that this scale stands as a characteristic scale of its statistics.

Direct inspection of SRI distributions does not reveal any other characteristic scale than x_c. While this would not suffice for concluding that x_c is the only characteristic scale involved in SRI distributions, it actually appears that this property is true. This may be understood by the following reasoning.

If there was no characteristic scale, scale-invariance would be fulfilled. Distributions would then follow relations (17, 18) in contrast with relations (32, 33). Some characteristic scales are thus in order. However, if two of us $x_{c,1}$ $x_{c,2}$ were to exist, their ratio $x_{c,1}/x_{c,2}$ would be also a characteristic of the distribution: the statistics would then not satisfy scale-ratio invariance. Hence, *one* characteristic scale and one *only* is involved in SRI distributions (32, 33). This is confirmed in Appendix A.2 by another method.

5.2 Sets of symmetric distributions

We denote $\mathcal{SI}$ and $\mathcal{SRI}$ the spaces of symmetric distributions with respect to scale-invariance and scale-ratio invariance. They correspond to the distributions (17, 18) and (32, 33):

$$\mathcal{SI} = \{p_e(.,.;.,.)\ ;\ e \in \mathcal{R}\}$$
$$\mathcal{SRI} = \{p_{f,x_c}(.,.;.,.); f \in \mathcal{R}\ ;\ x_c \in \mathcal{R}\}.$$

To emphasize the respective character of the two scale symmetries, we address below their effects on these sets. This leads us to compare the respective properties of these sets and their link with the SDL.

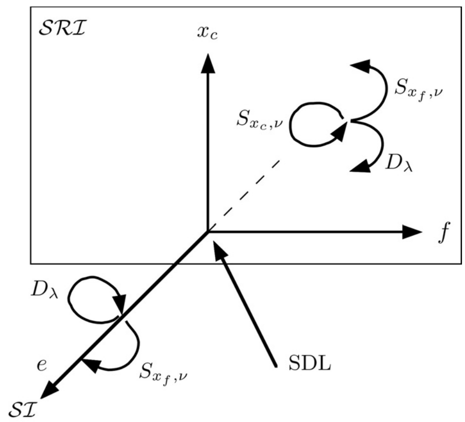

Fig. 7. Sketch of the sets of symmetric distributions $\mathcal{SI}$, $\mathcal{SRI}$, and of the orbits of their elements by dilation D_λ and stretch $S_{x_f,\nu}$. Both sets are invariant by dilation and stretch. However, the elements of $\mathcal{SI}$, the scale-invariant distributions, are invariant by any D_λ. Those of $\mathcal{SRI}$, the scale-ratio invariant, are invariant by any stretch $S_{x_c,\nu}$ where x_c labels the characteristic scale of these distributions. Notice that none of these sets is included in the other and that their intersection is reduced to the SDL.

5.2.1 Orbits by dilation

Applying a dilation D_λ to the distributions (17, 18, 32, 33) yields the following changes:

$$p_e(x_i, x_j; a, b) \to p_e(x_i, x_j; a, b)$$
$$p_{f,x_c}(x_i, x_j; a, b) \to p_{f,x_c/\lambda}(x_i, x_j; a, b).$$

This reveals the following properties that are displayed in Figure 7:

1. dilations preserve each SI distribution p_e;
2. dilations modify the SRI distributions p_{f,x_c};
3. dilations change one SRI distribution for another;
4. dilations conserve the exponents e or f of SI or SRI distributions;
5. dilations change the characteristic scales of SRI distributions: $x_c \to x_c/\lambda$.

The first property simply follows from scale-invariance. The second property shows an actual difference between SI and SRI which emphasizes the different natures of these scale symmetries. The third property shows that, although the set $\mathcal{SRI}$ is not invariant by dilation, it is *globally* conserved: $D_\lambda(\mathcal{SRI}) = \mathcal{SRI}$. This corroborates the fact that dilations do not introduce by themselves additional characteristic scale and thus any characteristic scale-ratio in distributions. This is confirmed by the fourth property which shows that exponents stand as invariants of the orbits generated by dilations. The changes undergone along an orbit of a SRI distribution is then stated by the last property: they affect the characteristic scale x_c. This stresses the fact that dilations actually change scales and, therefore, characteristic scales.

5.2.2 Orbits by stretch

Applying a stretch $S_{x_f,\nu}$ to the distributions (17, 18, 32, 33) yields the following changes:

$$p_e(x_i, x_j; a, b) \to p_{\nu e}(x_i, x_j; a, b)$$
$$p_{f,x_c}(x_i, x_j; a, b) \to p_{f,S^{-1}_{x_f,\nu}(x_c)}(x_i, x_j; a, b).$$

This reveals the following properties that are displayed in Figure 7:

1. stretch modifies each SI distribution p_e by changing it into another SI distribution;
2. stretch modifies each SRI distribution by changing it into another SRI distribution;
3. stretch changes the exponent of SI distributions: $e \to \nu e$;
4. stretch conserves the exponent of SRI distributions $f \to f$ but changes their characteristic scale: $x_c \to S^{-1}_{x_f,\nu}(x_c)$.

The first property shows that, although the set $\mathcal{SI}$ is not invariant by stretch, it is *globally* conserved: $S_{x_f,\nu}(\mathcal{SI}) = \mathcal{SI}$. This corroborates the fact that stretch does not introduce neither characteristic scale nor characteristic scale-ratio by itself. This is confirmed by the second property which shows that the set $\mathcal{SRI}$, although not invariant by stretch, is also *globally* conserved: $S_{x_f,\nu}(\mathcal{SRI}) = \mathcal{SRI}$. The third property indicates that a stretch changes the scale ratios, as expected from its definition. The last property shows that exponents f refer to a property common to different scale ratios. It also reveals the covariance property of SRI distributions: for a given change of scale ratio, i.e. for a given ν, SRI distributions can be invariant by the corresponding stretching transformation provided that their characteristic scale x_c stands as the *fixed point* x_f of this transformation: $x_f = x_c$.

5.2.3 Structures

Both sets of SI or SRI distributions are globally *stable* by dilation or stretch, but some of their distributions actually change under these transformations.

That no SI distribution change by dilation follows from their absence of characteristic scale. However, the fact that their exponent changes by stretch except when it is zero, $e = 0$, reveals that they involve a characteristic scale-ratio, except for the SDL. In an analogous way, the modification of SRI distributions by dilation, except for $f = 1$, shows that they involve a characteristic scale, except for the SDL. On the other hand, the fact that scale ratios can be changed (i.e. $\nu \neq 1$) while preserving SRI distribution (i.e. by $S_{x_c,\nu}(.)$) means that they involve no characteristic scale ratio. Their change under other stretching transformations, i.e. $S_{x_f,\nu}(x_c) \neq x_c$, can thus be attributed to the additional dilation $D_{(x_f/x_c)^{1-\nu}}$ involved in these transformations in comparison to the covariant one: $S_{x_f,\nu}(.) = D_{(x_f/x_c)^{1-\nu}} \circ S_{x_c,\nu}(.)$.

Following this analysis, the set $\mathcal{SI}$ of SI distributions is made of distributions involving *no* characteristic scale but a *single* characteristic scale-ratio, except for the SDL which involves *none*. Only the SDL is thus scale-ratio invariant in the $\mathcal{SI}$ set.

Similarly, the set $\mathcal{SRI}$ of SRI distributions is made of distributions involving *no* characteristic scale-ratio but a *single* characteristic scale, except for the SDL which involves *none*. Only the SDL is thus scale invariant in the $\mathcal{SRI}$ set.

The two scale symmetries are thus independent one of the other, neither of them being a consequence of the other:

$$\mathcal{SI} \neq \mathcal{SI} \cap \mathcal{SRI}; \quad \mathcal{SRI} \neq \mathcal{SRI} \cap \mathcal{SI} \qquad (57)$$

the SDL being the only distribution that satisfies both of them:

$$\mathcal{SI} \cap \mathcal{SRI} = \{\mathtt{SDL}\}. \qquad (58)$$

Interestingly, as shown in Section 4.2, this common distribution appears as an attractor of other distributions by dilation, stretch, and merging.

6 Conclusion

Since more than a century, an ubiquitous logarithmic distribution of data applying to many various numerical data sets has been discovered by Newcomb [1]: the "significant digit law". About fifty years later, it was supported by a wide investigation performed by Bendford on many everyday-life numerical data sets [2]. Since then, this observation has been used to detect frauds in business data [7,9] or to improve computer designs [5]. Yet, the reasons of its widespread occurrence has still remain the field of intense statistical and mathematical investigation [3,5,6,10–15,18,22].

We have shown here that the SDL results from *two* similar, albeit different, scale symmetries: the *scale* invariance and the *scale-ratio* invariance. For this, the relevance of these symmetries to large common data sets has first been legitimized from a phenomenological point of view. Then the distributions symmetric by change of scales or of scale-ratios have been determined from a criterion of *covariance*. The *only* distribution satisfying *both* symmetries has been found to be the SDL.

This specific distribution has been shown to possibly *attract* other distributions by dilation, stretch or merge of uncorrelated enough data. This is in particular the case when data are sampled in a fixed window depending on the local properties of the p.d.f of initial data. When the whole data sets are analyzed in an unbounded window, this remains true for stretch and merge but gets wrong for dilations.

Altogether, these results explain the ubiquity of this distribution and offer a deep insight into its nature. This can be used to clarify the situations in which the SDL might be relevant (e.g. finance, digital analysis) by investigating whether scale-invariance *and* scale-ratio invariance make sense for the corresponding systems or by determining the reasons of its possible failures (e.g. psychological barriers [22], fraud [6]).

Beyond this characterization of the SDL, the present analyses draw attention on a scale symmetry which is mostly ignored but which could be of paramount importance: the *scale-ratio* invariance. In particular, it has been found here as essential as the well-known scale-invariance for clarifying the nature of the SDL. Interestingly, it has proved to be more appropriate than the scale-invariance to investigate the differences between the concepts of invariance and of covariance. In view of the paramount generality of symmetries, we may guess that this invariance by change of scale-ratio might play a similar important role in other topics where multiple scales are involved [20,21]. Then, the present analyses would provide a useful guide for implementing its implications.

Appendix A

A.1 Selection of the SDL from the implications of statistical symmetries on the digit distribution

We consider a base b, a set of data $\{d\}$ and the distribution $D(x, y, b)$, $(x, y) \in [1, b[^2$, which expresses the probability of finding, in the base b, the mantissa $\sigma = db^{-n}$, $\sigma \in [1, b[$, $n \in \mathcal{Z}$, of data d in the range $[x, y[$.

We first assume that this statistics is covariant by dilation of data $d \to d' = \lambda d$, i.e. is scale-invariant, and we seek to identify the corresponding distributions. Following the analysis of Section 3.1.1, it is worth restricting the set of data to a bounded range $B_N = [b^{-N}, b^N[$, $N > 0$, in order to avoid the possibility of breaking of the concept of probability (see App. A.5). This precaution, which makes difference with the analysis of Pinckham [11], will prove to be essential for recovering all the sought distributions. Taking memory of it, we shall keep an explicit label for the data range B_N in the digit distribution by denoting this distribution $D(x, y, b, N)$.

We first notice that, for data belonging to the bounded range B_N and involving the most significant part σ of their mantissa in the sub-range $[x, y[$, $1 < x$, $y < b$, there exists a range of dilation factors $\lambda \in [1, \Lambda]$ small enough to avoid a jump of σ from the vicinity of b to 1. Then, dilation of data implies the *same* dilation of the most significant part of their mantissa, i.e.:

$$d \to d' = \lambda d \Longrightarrow \sigma \to \sigma' = \lambda\sigma. \qquad (A.1)$$

We shall restrict the following analysis to this case.

Let us denote $H(x, y, b)$ the histograms of the most significant parts of data mantissae in the range $[x, y[$ in the base b. We first notice that, by definition, the histograms of the initial distribution, H, and of the dilated distribution, H', are related one to the other by:

$$H'(x', y', b) = H(x, y, b). \qquad (A.2)$$

However, when statistics are scale-invariant, the repartition of the most significant parts of data mantissae in a range $[x, y[$ should not be able to reveal whether a scale dilation has been applied to data. This means that their histograms should actually be the same:

$$H'(x', y', b) = H(x', y', b) \quad \text{(A.3)}$$

so that:

$$\forall (x, y),\ 1 < x\ ,\ y < b\ ;\ \exists \Lambda\ ;\ \forall \lambda \in [1, \Lambda]$$
$$H(\lambda x, \lambda y, b) = H(x, y, b). \quad \text{(A.4)}$$

An essential point to realize, however, is that, because of dilation, the data contained in the fixed data range B_N before and after dilation are *not the same*, so that their number may actually *differ*. In particular, those contained in $[b^N/\lambda, b^N[$ (resp. $[b^{-N}/\lambda, b^{-N}[$) have been expelled from (resp. included in) B_N. For this reason, the relationship (A.4) between scale-invariant histograms only yields a proportionality between distributions up to an *unknown* prefactor $\mu(\lambda)$ corresponding to the relative change of the number of data:

$$\forall (x, y),\ 1 < x,\ y < b;\ \exists \Lambda;\ \forall \lambda \in [1, \Lambda]\ ;\ \exists \mu(\lambda);$$
$$D(\lambda x, \lambda y, b, N) = \mu\,(\lambda) D(x, y, b, N). \quad \text{(A.5)}$$

Expressing, as in section 3.1.3, the digit distribution with a mono-variate function $g(.)$:

$$D(x, y, b, N) = \frac{g(y) - g(x)}{g(b) - g(1)} \quad \text{(A.6)}$$

one obtains, from (A.5), the criterion (11). The solutions are thus not only the SDL but also the distributions (17). The existence of scale-invariant distributions other than the SDL traces back to the prefactor μ which actually differs from unity except for the SDL. This explains that overlooking it restricts the selection to the sole SDL [3,11].

We now assume that statistics of the most significant part of data mantissa are covariant by change of scale ratio, i.e. by transformations $S_{x_f,\nu}(.)$. Notice that data $d = \sigma b^n$ are transformed into $d' = S_{x_f,\nu}(\sigma) b^{\nu n} = \sigma' b^{n'}$ with $\sigma' \in [1, b[$, so that the link between σ and σ' depends on n, i.e. on the magnitude of data d. To avoid handling this parametrization, we restrict ourselves to $d = \sigma b$, σ belonging to the sub-range $[x, y[\subset [1, b[$. Then, for ν and x_f close enough to 1, the transformed value of σ keeps in between 1 and b, so that $\sigma' = x_f (\sigma/x_f)^\nu b^{\nu-1}$ (notice that the latter condition, x_f close enough to 1, is actually superfluous since the values of $x_f(\nu)$ that ensure invariance of distributions may be expected to be continuous with respect to ν and to satisfy $x_f(1) = 1$). Then, the same reasoning as for scale-invariance shows us that:

$$\forall (x, y),\ 1 < x\ ,\ y < b;\ \exists D\ \ ;\ \ \forall \nu \in [1, D]\ ;\ \exists x_f, \exists \mu(\lambda);$$
$$D(S_{x_f,\nu}(x), S_{x_f,\nu}(y), b, N) = \mu\,(\lambda) D(x, y, b, N). \quad \text{(A.7)}$$

Using invariance by dilation and considering log-variables (24), criterion (A.7) turns out to be equivalent to the scale-covariant criterion (26).

Therefore, as for distributions of data, the scale-invariance and the scale-ratio invariance of digit distributions altogether select the SDL.

A.2 SRI and characteristic scales

Consider a SRI distribution. By definition, it is invariant by stretching transformations $S_{x_f,\nu}(.)$ for any ν and suitably chosen $x_f(\nu)$. In particular, it is also invariant by the combination:

$$S^{-1}_{x_{f2},\nu_2}(.) \circ S^{-1}_{x_{f1},\nu_1}(.) \circ S_{x_{f2},\nu_2}(.) \circ S_{x_{f1},\nu_1}(.) = D_\lambda(.) \quad \text{(A.8)}$$

which is a power law with zero exponent, i.e. a dilation $D_\lambda(.)$. Its dilation factor λ:

$$\lambda = \left(\frac{x_{f2}}{x_{f1}}\right)^{(1-1/\nu_1)(1-1/\nu_2)} \quad \text{(A.9)}$$

differs from unity as soon as the characteristic scales (x_{f1}, x_{f2}) differ, $x_{f1} \neq x_{f2}$, and both the exponents ν_1, ν_2 are different than unity.

Assume now that there exists different values of the characteristic scales $x_f(\nu)$ for this SRI distribution. As distribution functions are continuous, the function $x_f(.)$ must be continuous too. The values of λ generated in (A.9) by varying the exponents ν_1, ν_2, would then extend continuously on a range including, but not restricted to, unity. However, by combination of dilations, all dilation factors could then be recovered. Accordingly, the SRI distributions would be preserved by any dilation: it would be scale-invariant.

Requiring more than one characteristic scale for SRI distributions thus implies that they involve none! This paradoxical conclusion shows that SRI distributions can only involve a *single* characteristic scale x_c: $\exists ! x_c; \forall \nu, x_f(\nu) = x_c$.

A.3 Short derivation of scale-ratio invariant solutions

Following Appendix A.2, there exists a single characteristic scale x_c for scale-covariant distributions. This implies that x_c must be the fixed point of the transformations $S_{x_f,\nu}(.)$ expressing the covariance, otherwise each of them would generate an infinite series of characteristic scales: $S^n_{x_f,\nu}(x_c)$. This in particular means that *all* the transformations yielding covariance have the *same* fixed point, x_c: $\forall \nu, x_f(\nu) = x_c$.

Apply now the dilation $D_{x_c^{-1}}(.)$. It changes probabilities p in $p \to \hat{p} = p \circ D_{x_c^{-1}}$ and brings the characteristic scale x_c to unity: $x_c \to 1$. As the characteristic scale of $\hat{p}$ is unity, the affine transformation $A_{M,\nu}(.)$ (22) then reduces to a dilation. The distributions covariant by $A_{M,\nu}(.)$ (25) then simply correspond to scale-invariant distributions (17, 18). Going back to the distributions p by using log-variables (24), one then recovers the scale-ratio invariant solutions (32, 33).

A.4 Formulation of covariances with mantissa and base

For *fixed* ξ and ρ, let us write number x in logarithmic representation: $x = \xi \rho^{\tilde{x}}$. This yields a bijection between x and $\tilde{x}$.

Dilation $D_\lambda(.)$ then corresponds to a dilation of ξ at fixed ρ and $\tilde{x}$:

$$D_\lambda \; : \; (\xi, \rho, \tilde{x}) \to (\lambda\xi, \rho, \tilde{x}). \tag{A.10}$$

Power law change $S_{\mu,\nu}(.)$ reduces to a stretch of ρ and a power law change of ξ at fixed $\tilde{x}$:

$$S_{x_f,\nu} \; : \; (\xi, \rho, \tilde{x}) \to (S_{x_f,\nu}(\xi), \rho^\nu, \tilde{x}). \tag{A.11}$$

In this framework, the covariance of SI distribution or SRI distribution can be simply checked:

– SI distributions $p_e(.,. \mid .,.)$ (17) read:

$$p_e(x_i, x_j; a, b) = \frac{\alpha^{\tilde{x_j}} - \alpha^{\tilde{x_i}}}{\alpha^{\tilde{x_b}} - \alpha^{\tilde{x_a}}} \tag{A.12}$$

with $\alpha = \rho^e$. The absence of ξ in this expression shows the invariance by change of scale units ξ, i.e. scale-invariance. Here covariance by dilation of p_e actually reduces to its *invariance* by dilation and thus to the absence of mantissa ξ.

– SRI distributions $p_{f,x_c}(.,. \mid .,.)$ (32) read:

$$p_{f,x_c}(x_i, x_j \mid a, b) = \frac{(A+\tilde{x_j})^f - (A+\tilde{x_i})^f}{(A+\tilde{x_b})^f - (A+\tilde{x_a})^f} \tag{A.13}$$

with $A = \ln[\xi/x_c]/\ln(\rho)$.
Covariance by power law change $S_{x_f,\nu}(.)$ thus reduces to the invariance, for any ν, of parameter A by at least one $S_{x_f,\nu}(.)$. This is indeed ensured when the scale units ξ is taken as the characteristic scale x_c, since then $A = 0$ whatever ρ. Power law changes $S_{\xi,\nu}(.)$ then gives covariance since ξ does not change and, therefore, $A = 0$ neither.
More generally, for $\xi \neq x_c$, invariance of A for a power law change $S_{x_f,\nu}(.)$ requires that ξ/x_c gets stretched as ρ: $\rho \to \rho^\nu$ and $\xi/x_c \to \xi'/x_c = (\xi/x_c)^\nu$, i.e. $\xi' = S_{x_c,\nu}(\xi)$. As by definition $\xi' = S_{x_f,\nu}(\xi)$, this means that covariance can only be obtained for power law changes $S_{x_f,\nu}(.)$ involving x_c as their fixed point: $x_f = x_c$.

A.5 Breaking of probability for the SDL

As feared in Section 3.1.1, the concept of probability breaks down when no restriction on the set of picked-up numbers is imposed, i.e. when either $a = 0$ or $b = \infty$. This arises from either a vanishing or a divergence of probability $p(x_i, x_j \mid a, b)$ for all conditional intervals $[x_i, x_j[$. In particular:

– Scale invariant distributions

$$\forall (x_i, x_j) \;,\; b \to \infty \Rightarrow p_e(x_i, x_j \mid a, b) \to 0$$
$$a \to 0 \Rightarrow p_0(x_i, x_j \mid a, b) \to 0.$$

For scale invariant distributions, the concept of probability breaks down on any infinite set $[a, \infty[$ and, for the distribution which is also scale-ratio invariant, on any finite set $[0, b[$ starting at zero.

– Scale-ratio invariant distributions

$$\forall (x_i, x_j) \;,\; b \to \infty \Rightarrow p_e(x_i, x_j \mid a, b) \to 0$$
$$a \to 0 \Rightarrow p_e(x_i, x_j \mid a, b) \to 0.$$

For scale-ratio invariant distributions, the concept of probability breaks down on any infinite set $[a, \infty[$ and on any finite set $[0, b[$ starting at zero.

This shows that the significant digit law is meaningful only if considered on sets involving finite non-zero bounds.

A.6 Solutions of the functional criterion for scale-invariance

We recall the functional criterion (11) for scale-invariance:

$$\forall\lambda, \; \exists h(.), \; \exists k(.);$$
$$\forall x, \; g(\lambda x) = h(\lambda) \; g(x) + k(\lambda)$$

where λ denotes the scale dilation factor.

Transitivity between dilations of scale factor λ_1 and λ_2 yields:

$$\begin{aligned} g(\lambda_1\lambda_2 x) &= h(\lambda_1) \; g(\lambda_2 x) + k(\lambda_1) \\ &= h(\lambda_1)h(\lambda_2)g(x) + h(\lambda_1)k(\lambda_2) + k(\lambda_1) \\ &= h(\lambda_2)h(\lambda_1)g(x) + h(\lambda_2)k(\lambda_1) + k(\lambda_2) \\ &= h(\lambda_1\lambda_2)g(x) + k(\lambda_1\lambda_2) \end{aligned}$$

so that:

$$h(\lambda_1\lambda_2) = h(\lambda_1)h(\lambda_2) \tag{A.14}$$

$$\begin{aligned} k(\lambda_1\lambda_2) &= h(\lambda_1)k(\lambda_2) + k(\lambda_1) \\ &= h(\lambda_2)k(\lambda_1) + k(\lambda_2). \end{aligned} \tag{A.15}$$

Relation (A.14) shows that function $H(.) = \ln[h(.)]$ is additive with respect to variable $\Lambda = \ln(\lambda)$:

$$H(\Lambda_1 + \Lambda_2) = H(\Lambda_1) + H(\Lambda_2). \tag{A.16}$$

This implies the linearity of $H(.)$ on the rational numbers: $\forall r \in \mathcal{Q}, \; \forall x, \; H(rx) = rH(x)$. This property shows that $H(.)$ is either continuous everywhere or discontinuous everywhere. Rejecting the pathological case of a discontinuity for all scale factors, we shall consider a linear $H(.)$. Back to function $h(.)$, this gives:

$$\exists e; \; \forall x, \; h(x) = x^e \; h(1) \tag{A.17}$$

and, according to (A.14), $h(1) = 1$ for non-zero $h(.)$.

Let us distinguish two cases:

– $e = 0$
For $e = 0$, function $h(.)$ is constant and relation (A.15) reads:

$$k(\lambda_1\lambda_2) = k(\lambda_1) + k(\lambda_2). \tag{A.18}$$

As for function $H(.)$ (A.16), this relation implies that $k(.)$ is linear in variable $\Lambda = \ln(\lambda)$ provided it is not everywhere discontinuous:

$$\exists C; \; \forall\lambda; \; k(\lambda) = C \; \ln(\lambda). \tag{A.19}$$

Altogether, relations (**??**) (A.17) (A.19) then yield:

$$\exists C;\ \forall x,\ \forall\lambda,\ g(\lambda x) = g(x) + C\ln(|\lambda|)$$
$$\exists(\alpha,\beta);\ \forall x,\ g(x) = \alpha\ \ln(|x|) + \beta. \qquad \text{(A.20)}$$

This corresponds to relation (15).

– $e \neq 0$

For $e \neq 0$, relation (A.15) yields:

$$k(\lambda_2)(1-\lambda_1^e) = k(\lambda_1)(1-\lambda_2^e) \qquad \text{(A.21)}$$
$$\exists C\ ;\ k(\lambda_1) = C\ (1-\lambda_1^e) \qquad \text{(A.22)}$$

so that, altogether, relations (**??**, A.17, A.21) give:

$$\exists C\ ;\ \forall x\ ,\ \forall\lambda\ ,\ g(\lambda x) - C = \lambda^e(g(x) - C)$$
$$\exists(\alpha,\beta)\ ;\ \forall x\ ,\ g(x) = \alpha x^e + \beta. \qquad \text{(A.23)}$$

This corresponds to relation (16).

A.7 Solutions of the functional criterion for scale-ratio invariance

We recall the functional criterion (26) for scale-ratio invariance:

$$\forall\nu,\ \exists M(.),\ \exists H(.), \exists K(.)\ ;$$
$$\forall X,\ G[\nu X + M(\nu)] = H(\nu)\ G(X) + K(\nu) \qquad \text{(A.24)}$$

where ν is the stretch factor of scale-ratios induced by power law transformations $S_{x_f,\nu}(.)$.

Transitivity between stretching of scale-ratio yields:

$$G[\nu_1\nu_2 X + M(\nu_1\nu_2)]$$
$$= H(\nu_1\nu_2)\ G(X) + K(\nu_1\nu_2)$$
$$= H(\nu_1)\ G[\nu_2 X + \frac{M(\nu_1\nu_2) - M(\nu_1)}{\nu_1}] + K(\nu_1)$$
$$= H(\nu_1)\ H(\nu_2)\ G\left[X + \frac{M(\nu_1\nu_2) - M(\nu_1) - \nu_1 M(\nu_2)}{\nu_1\nu_2}\right]$$
$$+ H(\nu_1)\ K(\nu_2) + K(\nu_1)$$

so that:

$$\exists(A,B,C);\ \forall X\ ,\ G(X) = AG(X+B) + C \qquad \text{(A.25)}$$

with

$$A = \frac{H(\nu_1)H(\nu_2)}{H(\nu_1\nu_2)}$$
$$B = \frac{M(\nu_1\nu_2) - M(\nu_1) - \nu_1 M(\nu_2)}{\nu_1\nu_2}$$
$$C = \frac{H(\nu_1)K(\nu_2) + K(\nu_1) - K(\nu_1\nu_2)}{H(\nu_1\nu_2)}.$$

Let us assume that A is not unity. The functional relation (A.25) then yields with $\exp(aB) = A$, $b = C/(A-1)$ and $\tilde{G}(X) = \exp(aX)[G(X) + b]$, the relationship:

$$\tilde{G}(X+B) = \tilde{G}(X) \qquad \text{(A.26)}$$

Denoting $P_B(X)$ this B-periodic function, we obtain:

$$\exists(a,b,B), \quad a = \ln(A)/B,\ b = C/(A-1);$$
$$G(X) = \exp(-aX)\ P_B(X) - b.$$

Then criterion (A.24) reads:

$$\forall\nu,\ \ \exists(a,b,B,M,H,K);\ \forall X,$$
$$\exp(-a\nu X)\ \exp(-aM)\ P_B(\nu X + M) - b$$
$$= H(\nu)\exp(-aX)P_B(X) - H(\nu)b + K(\nu). \qquad \text{(A.27)}$$

This can only be achieved for $a = 0$, i.e. $A = 1$, in contradiction with our assumption.

The ratio A is thus equal to unity: $A = 1$. Let us now assume that B is not zero. Then relation (A.25) yields with $\tilde{G}(X) = G(X) + aX$ and $a = C/B$ the relationship: $\tilde{G}(X) = \tilde{G}(X+B) = P_B(X)$. where $P_B(.)$ denotes a B-periodic function. This gives $G(X) = P_B(X) - aX$ which, together with criterion (A.24), implies:

$$\forall\nu,\ \exists(a,B,M,H,K);\ \forall X$$
$$P_B(\nu X + M) = HP_B(X) + a(\nu - H)X$$
$$+ aM + K. \qquad \text{(A.28)}$$

This can only be achieved for $a = 0$, i.e. $C = 0$, and $B = 0$. The later results follows from the fact that $P_B(\nu X + M)$ is B/ν-periodic in X, so that the function $P_B(.)$ is actually B/ν^p-periodic, $p \in \mathcal{Z}$. Being assumed to be continuous, it can then only be a constant: $B = 0$, $P_B(.) = K$.

Let us compute B by applying the scale transformations indexed 1 and 2 in either order. Commutativity of this operation combined with the fact that in either case $B = 0$, yields:

$$\forall(\nu_1,\nu_2),\ M(\nu_1)(1-\nu_2) = M(\nu_2)(1-\nu_1) \qquad \text{(A.29)}$$

and:

$$\forall\nu,\ \exists D;\ M(\nu) = D(1-\nu). \qquad \text{(A.30)}$$

Defining $Y = X - D$, $\hat{G}(Y) = G(X)$, relation (A.30) combined with criterion (A.24) yields:

$$\forall\nu,\ \exists H(.),\ \exists K(.)\ ;$$
$$\forall Y,\ \hat{G}(\nu Y) = H(\nu)\ \hat{G}(Y) + K(\nu).$$

This is actually similar to criterion (**??**) whose solutions are given in (A.20, A.23).

Back to function $G(.)$ and variable x with $X = \ln(|x|)$, one then obtains, with $D = \ln(|x_f|)$:

$$\exists(\alpha,\beta,f,x_f)\,;$$
$$f \neq 0\,,\ g(x) = \alpha\ (\ln(|x/x_f|)^f + \beta$$
$$f = 0\,,\ g(x) = \alpha\ln(|\ln(|x/x_f|)|) + \beta. \qquad \text{(A.31)}$$

This corresponds to relations (30, 31).

References

1. S. Newcomb, Am. J. Math. **4**, 39 (1881)
2. F. Benford, Proc. Amer. Philos. Soc. **78**, 551 (1938)
3. T. Hill, Stat. Sci. **10**, 354 (1995); T. Hill, Proc. Amer. Math. Soc. **123**, 887 (1995); T. Hill, Amer. Math. Monthly **102**, 322 (1976)
4. Eur. Phys. J. B **22**, number 1, 2, 3 and 4, (2001)
5. P. Schatte, J. Inform. Process. Cybernet. **24**, 433 (1988)
6. Ph.D. Drake, M.J. Nigrini, J. of Acc. Ed. **18**, 127 (2000)
7. M.J. Nigrini, J. Amer. Taxation Assoc. **18**, 72 (1996)
8. C.L. Geiger, P.P. Williamson, Communication in Statistics, Simulation and Computation **33**, 229 (2004)
9. E. Ley, Amer. Stat. **50**, 311 (1996)
10. R. Raimi, Sci. Amer. **221**, 109 (1969); R. Raimi, Amer. Math. Monthly **83**, 521 (1976)
11. R. Pinckham, Ann. Math. Statist. **32**, 1223 (1961)
12. P. Diaconis, Ann. Probab. **5**, 72 (1977)
13. A. Adhikari, B. Sarkar, Sankhya Ser. B **30**, 47 (1968)
14. L. Pietronero, E. Tosatti, V. Tosatti, A. Vespigliani, Physica A **293**, 297 (2001)
15. M.A. Snyder, J.H. Curry, A.M. Dougherty, Phys. Rev. E **64**, 026222 (2001)
16. C.R. Tolle, J.L. Budzien, R.A. LaViolette, Chaos **10**, 331 (2000)
17. D. Knuth, *The Art of Computer Programming* (Addisson-Wesley, Reading, MA, 1969), Vol. 2, p. 219
18. B. Buck, A. Merchant, S. Perez, Eur. J. Phys. **14**, 59 (1993)
19. W. Feller, *An introduction to probability Theory and its Applications* (Wiley and Sons, 1966)
20. A. Pocheau, in *Mixing, Chaos and Turbulence*, edited by H. Chaté, J.-M. Chomaz, E. Villermaux Nato ASI Series B (Physics), Vol. 373 (Plenum Press, 2000) pp. 187–204; A. Pocheau, Phys. Rev. E **49**, 1109 (1994); A. Pocheau, D. Queiros-Condé, Phys. Rev. Lett. **76**, 3352 (1996)
21. A. Pocheau, Europhys. Lett. **43**, 410 (1998)
22. M.J.K. de Ceuster, G. Dhaene, T. Shatteman, J. Empirical Finance **5**, 263 (1998)