

1 Generalities

1.1 Information-Theoretic Maximum Entropy Principle and Inverse Problem

1.1.1 Information-Theoretic Maximum Entropy Principle

According to recent literature (Golan, Judge, & Miller, 1996; Golan, 2008), the information-theoretic maximum entropy principle is a coincident junction of two lines of research: inferential statistics and statistical thermodynamics.

The first line of research emerged in the beginning of the 18th century through the work of Bernoulli (Jaynes, 1957), (Halmos & Savage, 1949), Bayes (1763), and Laplace (1774). They developed the Principle of Sufficient Reason, which consists of determining the state of the system on the basis of limited information (moments) from a subsystem. This principle was later extended in the last century by Jeffreys (1946), Cox (1946), and Jaynes (1957b) to the principle of “not telling more than you know,” thus suggesting the necessity of avoiding additional hypotheses imposed merely to simplify the problem to be solved. The purpose of all of the above authors’ research was to retrieve characteristics of a general population on the basis of limited information from a possibly non-representative sample of that population, out of risky or non-convenient hypotheses.

The second line of research is represented, amongst others, by Maxwell (1871), Boltzmann (1871), Cauchy (1855), Weierstrass (1886), Lévy and Gibbs (Gibbs, 1902), Shannon (1948), Jaynes (1957, 1957b), Rényi (1961), Bregman (1967), Mandelbrot (1967), Tsallis (1988). Its main objective was to provide mathematical formalism to statistical modelling of physical information related to natural phenomena. Thanks to the celebrated work of Tsallis (1988), on non-extensive thermodynamics², this second line elegantly extended its multidisciplinary applications to “auto-organized systems” and to the social sciences, particularly in financial fields.

The ascent and development of the post-war information theory-based, maximum entropy proposed by Shannon (1948) can be viewed as a major step toward the rapid extension of the discipline. Less than a decade was needed to develop the information-theoretic principles of statistical inference, inverse problem solution methodology based on Gibbs-Shannon maximum entropy, and its generalizations by Kullback and Leibler (1951), Kullback (1959) and Jaynes (1957b). The above authors developed, in particular, fundamental notions in statistics, such as sufficiency and efficiency

² Currently, this theory—undoubtedly the best—generalizes Boltzmann-Gibbs statistics for describing the case of anomalous systems characterized by non-ergodicity or metastable states. It thus better fits dynamic correlation of complex systems and can be better explained (e.g. Douglas, 2006), amongst many others.

(Halmos & Savage, 1979), a generalization of Cramer-Rao inequality (e.g., Kullback, 1959) and the introduction of a general linear model as a consistency restriction (Heckelei et al., 2008) through Bayesian philosophy. Thus, it became possible to unify heterogeneous statistical procedures via the concepts of information theory. Lindley (2008), on the other hand, had provided the interpretation that a statistical sample could be viewed as a noisy channel (Shannon's terminology) that conveys a message about a parameter (or a set of parameters) with a certain prior distribution. This new interpretation extended application of Shannon's ideas to statistical theory by referring to the information in a statistical sample rather than in a message.

Over the last two decades the literature concerned with applying entropy in social science has grown considerably and deserves closer attention. On one side, Shannon-Jaynes-Kullback-Leibler-based approaches are currently used for modelling economic phenomena competitively with classical econometrics. A new paradigm in econometrical modelling is taking place and finds its roots in the influential work of Golan, Judge, and Miller (1996). The present monograph constitutes an illustration of this.

As mentioned above, this approach is particularly useful in the case of solving inverse problems or ill-behaved matrices when we try to estimate parameters of an econometric model on the basis of insufficient information from an observed sample, and this estimation may concern the behaviour of an individual element within the system.

Insufficient information implies that we are trying to solve an ill-posed problem, which plausibly can arise in the following cases:

- data from sampling design are not sufficient and/or complete due to technical or financial limitations—*small area official statistics* could illustrate this situation;
- non-stationary or non-co-integrating variables are resulting from bad model specification;
- data from the statistical sample are linearly dependent or collinear for various reasons;
- Gaussian properties of random disturbance are put into question due to, amongst many others things³, systematic errors from the survey process;
- the model is not linear and *approximate* linearization remains the last possibility;
- aggregated (in time or space) data observations hide a very complex system represented, for instance, by a PL distribution, and multi-fractal properties of the system may exist.

³ It is not excluded that distribution law may be erroneously applied since, for instance, randomness is dependent on the experimental setup or the sophistication of the apparatus involved in measuring the phenomenon (Smith, 2001).

Using the traditional econometrical approaches in one or more of the above cases—without additional simplifying hypotheses—could lead to various estimation problems owing to the nonexistence of a bounded solution or the instability of estimates. Consequently, outputs from traditional econometrical approaches will display, at best, poor informative parameters. In the literature, there are other well-known techniques to cope with inverse problems or ill-conditioned data. Among them, two popular techniques deserve our attention: the bi-proportional RAS approach (and its variants), particularly used for updating or forecasting input/output matrices (Parikh, 1979) and the Moore-Penrose pseudo-inverse technique, useful for inverting irregular matrices (e.g., Green, 2003, p. 833). In spite of their popularity, both techniques present serious drawbacks in empirical investigations. In fact, the RAS techniques, in spite of their divergence information nature, remain less adapted to solving stochastic problems or to optimizing the information criterion function under a larger number of different prior constraining data. Since Moore-Penrose generalized inverse ensures a minimum distance ($Y-BX$) only when the matrix B has full rank, it will not reflect an optimal solution in other cases. Golan et al. (1996) have clearly shown higher efficiency of Shannon maximum entropy econometrics over the above cited methods in recovering unknown information when data or model design is poorly conditioned. The suggested superiority stands on the fact that it combines and generalizes maximum entropy philosophy (as in the second law of thermodynamics) and statistical theory of information attributes as a Bayesian information processing rule. As demonstrated convincingly by Golan (1996, 2006), Shannon entropy econometrics formalism may generalize least squares (LS) and the maximum likelihood (ML) approaches and belongs to the class of Bayesian method of moments (BMOM). It is worthwhile to point out that in the coming chapters many cases of cross-entropy (or minimum entropy) formalism will be used in place of maximum entropy. This is because, in this study, many problems to be treated involve information measuring in the context of the Kullback-Leibler framework.

This monograph does not intend to treat the case of high frequency series for which a rich literature already exists. We invite readers interested in the case of high frequency series to see, for instance, J.W. Kantelhardt (2008) for testing for the existence of fractal or multi-fractal properties, suggesting the case of a PL distribution.

1.2 Motivation of the Work

1.2.1 Frequent Limitations of Shannon-Gibbs Maximum Entropy Econometrics

In spite of a growing interest in the research community, some incisive critics have come forward to address Shannon-based entropy econometrics (e.g., Heckeles et al., 2008). According to some authors, generalized maximum entropy (GME) or cross-

entropy (GCE) econometrical techniques face at least three difficulties. The first is related to the specification and interpretation of prior information, imposed via the use of discrete support points, and assigning prior probabilities to them. The authors argue that there are complications that result from the combination of priors and their interaction with the criterion of maximum entropy or minimum cross-entropy in determining the final estimated *a posteriori* probabilities on the support space. The second group of criticisms questions the sense of the entropy objective function once combined with the prior and data information. The last problem, according to the same authors, refers to computational difficulties owing to the mathematical complexity of the model with an unnecessarily large number of parameters or variables.

Concerning the first criticism, the problem—selecting a prior support space and prior probabilities on it—exists since estimation outputs seem to be extremely sensitive to initial conditions. However, when there is a theory or some knowledge about the space on which parameters are supposed to be staying, the problem becomes tractable. In particular, when we have to estimate parameters in the form of ratios, the performance of entropy formalism is high. To this counterargument, it is worthwhile to add that GME or GCE formalism constitutes an approach based on the Bayesian efficient processing rule and, as such, prior values are not fixed constraints of the model; they combine and adapt with respect to other sets of information (e.g., consistency function) added to the model to update a new parameter level in the entropy criterion function.

The second problem concerns questioning the sense or interpretability of output probabilities from the maximum entropy criterion function once combined with real world probability-related restrictions. One cannot comment on this problem without making reference to the important contribution of Jaynes (1957, 1957b), who proposed a way to estimate unknown probabilities of a discrete system in the presence of less data point observations than parameters to be estimated through the celebrated example of Jaynes dice. Given a set of all possible ways of distribution resulting from all micro-elements of a system, Jaynes proposed using the one that generates the most “uncertain”⁴ distribution. To understand this problem, the question becomes a matter of combining philosophical interpretation of the maximum entropy principle with that of Jaynes’ formulation in the context of Shannon entropy. Depending on the type of entropy⁵ considered, output estimates will have slightly different meaning. However, all interpretations refer to parameter values that assure a long-run, steady-

⁴ Here we are in the realm of the second law of thermodynamics, which stipulates, in terms of entropy, that natural equilibrium of any set of events is reached once disorder inside them becomes optimal. This results from their property of having equal (ergodic system) odds to occur. In that state, we reach the maximum uncertainty about which event should occur in the next trial.

⁵ Later, for comparison, properties of the most well-known types of entropy in the literature will be presented.

state equilibrium of the system (relations defined by the model) with respect to data and other knowledge at hand, usually in the form of moments and/or normalization conditions. Owing to maximum entropy alone, the more consistent moments are or the more other *a priori* information binds, the more output probabilities will differ from those in a uniform distribution. Considering the above, interpretation of the maximum entropy model is far removed from interpretation of the classical model, especially in the case of the econometric linear model where estimates mean a change in the endogenous variable due to unitary change in an explicative variable, that is, in *ceteris paribus* conditions.

The last criticisms concern the burden arising from the computational and numerical process—a problem common to all complex, nonlinear systems. Thanks to recent developments of computer software, this problem is now less important.

In many empirical studies that attempt to solve inverse problems, the Shannon entropy-based approach is relatively efficient in recovering information. However, gaining in parameter precision requires good design of the prior. In particular, the point support space must fit into the space of the true population parameter values. As Golan et al. (1996) have shown, when prior design is weak, outputs of Shannon entropy econometrics will produce approximately the same parameter precision as traditional econometrical methods, such as LS or the ML, which means Shannon entropy could discount information not fitting the maximum entropy principle as expected.

The above criticisms of the Shannon entropy econometrics model remain relatively weak as has been shown through the preceding discussion.

According to us, the main drawback related to that form of model is due to the analytical function of constraining moments. In fact, as already suggested, long-range correlation and observed time invariant scale structure of high frequency series may still be conserved—in some classes of non-linear models—through a time—or space—aggregation process of statistical data. This raises the question of why this study proposes a new approach of Tsallis non-extensive entropy econometrics.

The next section provides a first answer by showing potential theoretical and then empirical drawbacks of the Shannon-Gibbs entropy model and potential advantages from the PL-related Tsallis non-extensive entropy approach.

1.2.2 Rationale of PL-Related Tsallis Entropy Econometrics and Low Frequency Series

This section presents the essence of the scientific contribution of this monograph to econometric modelling. For a few decades, PL has confirmed its central role in describing a large array of systems, natural and manmade. While most scientific fields have integrated this new element into their analytical approaches, econometrics and hence, economics globally, is still dwelling—probably for practical reasons—

on the Gaussian fundamentals. This study takes a step forward by introducing Tsallis non-extensive entropy to low frequency series econometric modelling. The potential advantages of this new approach will be presented, in particular, its capacity to analytically solve complex PL-related functions. Since any mathematical function form can be represented by a PL formulation, the importance of the proposed approach becomes clear. To be concrete, one of the complex nonlinear models is the fractionally integrated moving average (ARFIMA) model, which, to our knowledge, has remained non-tractable using traditional statistical instruments. An empirical application to solve such a class of models will be implemented at the end of Part V of this book.

According to several studies (Bottazzi & et al, 2007), (Champernowne, 1953), (Gabaix, 2008), a large array of economic laws take the form of a PL, in particular macroeconomic scaling laws, distribution of income, wealth, size of cities and firms⁶, and distribution of financial variables such as returns and trading volume. Ormerod and Mounfield (2012) underscore a PL distribution of business cycle duration. Stanley et al. (1998) have studied the dynamics of a general system composed of interacting units, each with a complex internal structure comprising many subunits, where the subunits grow in a multiplicative way over a period of twenty years. They found that this system followed a PL distribution. It is worthwhile to note the similarity of such a system with the internal mechanism of national accounts tables, such as a SAM, also composed of interacting economic sectors, each with a complex internal structure defined by firms exercising similar business. Ikeda and Souma (2008) have made an international comparison of labour productivity distribution for manufacturing and non-manufacturing firms. A PL distribution in terms of firms and sector productivity was found in US and Japanese data. Testing the Gibrat's law of proportionate effect, Fujiwara et al. (2004) have found, among others things, that the upper-tail of the distribution of firm size can be fitted with a PL (Pareto-Zipf law). The list of PL evidence here is limited to social science.

Since this study focuses on the immense potentiality of PL-related economic models, PL ubiquity in the social sciences will be underscored and a theorem showing the PL character of national accounts in its aggregate form will be presented.

In line with the rationale for the proposed methodology detailed below, the following from recent literature is evidence of entropy:

- Non-extensive entropy, as such, models the non-ergodic systems which compound Levy⁷ instable phenomena⁸ converging in the long range to the Gaussian basin of attraction. In the limiting case, non-extensive entropy converges to Shannon Gibbs entropy.

⁶ See (Bottazzi & et al, 2007) for different standpoints on the subject.

⁷ Shlesinger (Shlesinger, Zaslavsky, & Klafter, Strange Kinetics, 1993)

⁸ (Shlesinger & et al, Lévy Flights and Related Topics in Physics, 1995).

- PL-related Tsallis entropy should remain, even in the case of a low frequency series, a precious device for econometric modelling since the outputs provided by the exponential family law (e.g., the Gibbs-Shannon entropy approach) correspond to the Tsallis entropy limiting case when the Tsallis- q parameter equals unity.
- A number of complex phenomena involve long-range correlations which can be seen particularly when data are time scale-aggregated (Drożdż & Kwapień, 2012), (Rak & et al, 2007). This is probably because of the interaction between the functional relationships describing the involved phenomena and the inheritance properties of a PL or because of their nonlinearity. Delimiting the threshold values for a PL transition towards the Gaussian structure (or to the exponential family law) as a function of the data frequency amplitude is difficult since each phenomenon may display its own rate of convergence—if any—towards the central theorem limit attractor.
- Systematic errors from statistical data collecting and processing may generate a kind of tail queue distribution. Thus, a systematic application of the Shannon-Gibbs entropy approach in the above cases—even on the basis of annual data—could be misleading. In the best case, it can lead to unstable solutions.
- On the other hand, since non-extensive Tsallis entropy generalizes the exponential family law (Nielsen & Nock, 2012), the Tsallis- q entropy methodology fits well with high or low frequency series.

In the class of a few types of entropy displaying higher-order entropy estimators able to generalize the Gaussian law, Tsallis non-extensive entropy has the valuable quality of concavity—and then stability—along the existence interval characterizing most real world phenomena. As far as the q -generalization of the Kullback-Leibler (K-L) relative entropy index is concerned, it conserves the same basic properties as the standard K-L entropy and can be used for the same purpose (Tsallis, 2009).

The above-enumerated points imply that in cases where the assumed Levy law complexity is not verified by empirical observation, outputs from the non-extensive entropy model converge with those derived from Shannon entropy. In other words, errors which involve taking a sample as if it were PL-driven has no consequence on outputs if the truth model belongs to the Gaussian basin of attraction. This explains why in most empirical applications—but by no means all—both forms of entropy provide similar results and the entropic Tsallis- q complexity parameter then tends to converge to unity, revealing the case of a normal distribution. Empirical examples will be presented at the end of this document, and the strength of Tsallis maximum entropy econometrics will be demonstrated in different contexts.

In summary, the following are entropy function regularities:

- The Tsallis entropy model generalizes the Shannon-Gibbs model, which constitutes a converging case of the former for the Tsallis- q parameter equal unity.

- The Shannon-Gibbs model fits natural or social phenomena displaying Gaussian properties.
- PL high frequency time (space) series scaling—aggregating—does not always lead to Gaussian low frequency time (space) series. Additionally, the rate of convergence from the PL to the Gaussian model, if any, varies according to the form of the function used.

Is it judicious to replace Shannon-Gibbs entropy modelling by Tsallis non-extensive entropy for empirical applications?

The answer is yes, and this is the motivation for this study. There are at least three expected advantages to introducing Tsallis non-extensive econometric modelling:

1. A data generating system characterized by a low—or no—convergence rate from PL to Gaussian distribution only becomes analytically tractable when using Tsallis entropy formalism. (This will be proven through an econometrical model with constant substitution elasticity and then considered as an inverse problem to be estimated later.)
2. The Tsallis entropy model displays higher stability than the Shannon-Gibbs, particularly when systematic errors affect statistical data.
3. The Tsallis- q parameter presents an expected advantage of monitoring complexity of systems by measuring how far a given random phenomenon is from the Gaussian benchmark. In addition to other advantages, this can help draw attention to the quality of collected data or the distribution involved.

The choice of national accounts-related models for testing the new approach of non-extensive entropy econometrics is motivated by the empirical inability of national systems of economic information to provide consistent data according to macroeconomic general equilibrium. As a result, national account tables are generally not balanced unless additional—often contradictory—assumptions are applied to balance them. However, following the principle of not adding (to a hypothetical truth) more than we know, it remains preferable to deal with an *unbalanced* national accounts table. Trying to balance such a table implies that we are faced with ill-behaved inverse problems. According to the existing literature, and as will be seen through this monograph, entropy formalism remains the best approach to solving such a category of complex problems. The superiority of Tsallis non-extensive entropy econometrics over other known econometrical or statistical procedures results from its capacity to generalize a large category of most known laws, including Gaussian distribution.

1.3 National Accounts-Related Models and the Scope of this Work

Under the high frequency series hypothesis, we postulate that social and economic activities are characterized by complex behavioural interactions between socio-economic agents and/or economic sectors. Recent, *Big Data* for Official Statistics may illustrate such a complexity. This could mean that the supposed extreme events may appear systematically more (or less) frequently than expected (Gaussian scheme), implying internal and aggregated long-range correlation (over time, space, or both). The maximum entropy principle is best suited to estimating ill-behaved inverse problems and, in particular, models with ratios or elasticity as parameters. In this latter case, as we will see later, the support space area for unknown parameters coincides with the probability area over the space from zero to unity. Fortunately enough, due to its macroeconomic consistency, national account table structure reflects this property. In empirical macroeconomic investigations, the national accounts system of information plays a crucial role for modelling as it guarantees internal coherence of macroeconomic relations. Numerical information is embodied inside comprehensive statistical tables or balance sheets displaying algebraic properties of a matrix. Having in mind an economic or statistical inference investigation, mathematical treatment of information compounded inside these matrices is carried out by economists or statisticians on the basis of a priori information at hand. When such matrices are algebraically regular, traditional inverse methods can be applied to solve the problem of, for instance, estimating parameters that define relationships between the endogenous variable and its covariates. Nevertheless, in the social sciences, causality relationships linking both variables seldom have a one-to-one correspondence. In many cases, two or more different inputs or causes can lead to the same output or effect. Such different causal concomitances for the same output render the social or economic model indeterminate. In such cases, the recovery of a data generating system from the observed finite sample becomes impossible using the traditional statistical or econometric devices, such as the standard maximum likelihood method or the generalized method of moments. On mathematical grounds, this may result from an insufficient number of model data points with respect to the number of parameters to estimate. Such a sample is said to be ill-behaved. This situation leads to the lack of an optimal solution sought. Collinear variables, inadequate size of a small sample, or the poor quality of statistical data may lead to the same difficulties. Finally, taking into account the above deficiencies and anomalies, modellers have to deal with ill-behaved inverse problems most of the time. Following what has been said above, this monograph targets developing a robust approach generalizing Kullback-Leibler-Shannon entropy for solving inverse problems related to national account models in a way that reflects the complex relationships between economic institutions and/or agents. Statistical data from such complex interrelations are usually difficult to collect, incomplete, and defective. Additionally—and this may be one of the most important points—modelling national account table-related information involves

some class of nonlinear functions, otherwise only solvable using the PL model; thus, non-ergodic situations are involved.

The next area of national accounts modelling to be treated in this monograph is:

- Updating an input/output table when the problem is posed as inverse, with the possibility of adding extra sample information to the model in the form of an a priori and without any additional assumption;
- Forecasting an input/output table or its extended forms, such as the social accounting matrix (SAM), solely on the basis of yearly published national accounts concerning sectorial elements of final demand and gross domestic product;
- Deriving backward or forward multiplier coefficient impact on the basis of insufficient pieces of information;
- Demonstrating a method to forecast a sectorial energy final demand and total pollutants emission by production on the basis of an environmentally extended input/output table when basic information is missing;
- Presenting a computable general equilibrium model using the maximum entropy approach instead of calibration techniques to derive the parameters of CES functions,
- Estimating other nonlinear economic functions as inverse problems and conducting Monte Carlo experiments to test Tsallis entropy econometrics outputs;
- Presenting in detail, across different chapters, national account-related general equilibrium models before coming back to inverse problem solution techniques as suggested above.

The reader should be enriched not only by techniques for solving complex inverse problems but also by a thorough examination of different aspects of national account updating and modelling in the Walrasian spirit. To render the models presented here more consistent, emergent elements on an environmentally extended system of accounts will be included along with their impact on the general equilibrium framework and the optimum Pareto or social welfare.

Bibliography – Part I

- Bayes, T. (1763). An essay towards solving a Problem in the Doctrine of Chances, *Phil. Trans.* 53, 370–418 (1763). <http://rstl.royalsocietypublishing.org/content/53/370>
- Bottazzi, G., & et al. (2007). Invariances and Diversities in the Patterns of Industrial Evolution: Some Evidence from Italian Manufacturing Industries. *Small Business Economics* 29, pp. 137–159.
- Bregman, L.M. (1967). The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *Computational Mathematics and Mathematical Physics* 7(3), pp. 200–217.
- Champernowne, D.G. (1953). A Model of Income Distribution. *The Economic Journal* 63(250), pp. 318–351.
- Cox, R.T. (1946). Probability, Frequency, and Reasonable Expectation. *Am. Jour. Phys.*, 14, pp. 1–13.

- Douglas P. et al. (2006). Tunable Tsallis Distributions in Dissipative Optical Lattices. *Phys. Rev. Lett.* 96, 110601.
- Drożdż, S., & Kwapień, J. (2012). Physical approach to complex systems. *Physica reports*, pp. 115–226.
- Fujiwara, & et al. (2004). Gibrat and Pareto–Zipf revisited with European firms. , *Physica A* 344, 1–2, pp. 112–116.
- Gabaix, X. (2008, September). *Power Laws in Economics and Finance*. Retrieved from NBER: <http://www.nber.org/papers/w14299>
- Gibbs, J.W. (1902). *Elementary principles in statistical mechanics*. New York, USA: C. Scribner's Sons incl.
- Golan A. (2006), Information and Entropy Econometrics — A Review and Synthesis, *Foundations and Trends in Econometrics*, vol 2, no 1–2, pp 1–145.
- Golan, A., Judge, G., & Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. England: Wiley in Chichester.
- Green, W. (2003). *Basics of Econometrics, 5th edition*. NY: Prentice Hall.
- Halmos, P.R., & Savage, L.J. (1949). Application of the Radon-Nikodyma theorem to the theory of sufficient statistics. *Annals of Math. Stat.* 20, pp. 225–241.
- Hartley, R.L. (1928, July). Transmission of Information. *Bell System Technical Journal*.
- Heckelei, T. et al., (2008). *Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models*. Bonn: 2008/2, University of Bonn.
- Ikeda, Y., & Souma, W. (2009). *International comparison of Labor Productivity Distribution for Manufacturing and Non-manufacturing Firms*. Progress of Theoretical Physics Supplement, 179, pp. 93, Oxford University Press.
- Jaynes, E.T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, pp. 620–630.
- Jaynes, E.T. (1957b). *Probability Theory: The Logic Of Science*. USA: Washington University.
- Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. (R.S. London, Ed.) *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 186 (1007), pp. 453–461.
- Kantelhardt, J.W. (2008). *Fractal and multifractal times series*. Wisconsin, USA: Institute of Physics, Martin Luter University.
- Kullback, S., & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22, pp. 79–86.
- Kullback, S. (1959). *Information theory and statistics*. NY: John Wiley and Sons.
- Laplace, S.P. (1774). Memoir on the probability of causes of events. (English translation by S.M. Stigler, Ed.) *Statist. Sci.* 1(19), *Mémoires de Mathématique et de Physique, Tome Sixième*, 364–378.
- Lindley, D. (2008). *Uncertainty: Einstein, Heisenberg, Bohr, and the Struggle for the Soul of Science*. NY: Anchor.
- Mandelbrot, B. (1967). How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, 156(3775), pp. 636–638.
- Maxwell, J.C. (1871). *Theory of Heat*. New York: Dover.
- Nielsen, F., & Nock, R. (2012). A closed-form expression for the Sharma-Mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical* 45, p. 3.
- Ormerod, P., & Mounfield, C. (2012). Power law distribution of the duration and magnitude of recessions in capitalist economies: Breakdown of scaling. *Physica A*293, pp. 573–582.
- Parikh, A. (1979). Forecasts of Input-Output Matrices Using the R.A.S. Method. *The Review of Economics and Statistics* 61, 3, pp. 477–481.
- Rak, R., & et al. (2007). Nonextensive statistical features of the Polish stock market fluctuations. *Physica A* 374, pp. 315–324.

- Rényi, A. (1961). On Measures of Entropy and Information. In Univ. of California (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1 (pp. 547–561). Berkeley, Calif.: University of California Press.
- Kullback, S. (1959). *Information theory and statistics*, NY: John Wiley and Sons.
- Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, pp. 379–423 & 623–656.
- Shlesinger, M., & et al. (1995). *Lévy Flights and Related Topics in Physics*. Berlin: Springer.
- Shlesinger, M., Zaslavsky, & Klafter. (1993). Strange Kinetics. *Nature* 363, p. 31–37.
- Smith, J.D. (2001). Some Observations on the Concepts of Information-Theoretic Entropy and Randomness. *Entropy* 3[1], pp. 1–11.
- Stanley, H., & et al. (1998). Power Law Scaling for a System of Interacting Units with Complex Internal Structure. *Phys Rev Lett* 80, p. 7.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52, pp. 479–487.
- Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*. Berlin: Springer.
- Weierstrass, K. (1886). *Abhandlungen aus der Functionenlehre [Treatises from the Theory of Functions]*. Berlin, Germany: Julius Springer.

