About the parabolic relation existing between the skewness and the kurtosis in time series of experimental data

F Sattin, M Agostini, R Cavazzana, G Serianni, P Scarin and N Vianello

Consorzio RFX, Associazione EURATOM-ENEA sulla fusione, Corso Stati Uniti 4, Padova, Italy
E-mail: fabio.sattin@igi.cnr.it

Received 10 November 2008
Accepted for publication 12 February 2009
Published 31 March 2009
Online at stacks.iop.org/PhysScr/79/045006

Abstract
In this work, we investigate the origin of the parabolic relation between skewness and kurtosis often encountered in the analysis of experimental time series. We argue that the numerical values of the coefficients of the curve may provide information about the specific physics of the system studied, whereas the analytical curve per se is a fairly general consequence of a few constraints expected to hold for most systems.

PACS numbers: 06.30.−k, 07.05.Kf, 07.05.Rm

((Some figures in this article are in colour only in the electronic version.)
(SS) [6], originally conceived only to deal with fluctuations of the sea surface temperature, and showed that it worked equally well with the plasma density fluctuations reported in [1]: that is, two completely different problems may be modelled starting from the same formalism, a hint of some common underlying physics. We believe, however, that there is an even more general rationale and in this work we are going to propose reasons for this belief. Several scattered considerations, both of physical purely mathematical nature, will be assembled together. Their collection provides a consistent body of evidence in favour of our thesis, although—of course—a full-fledged proof cannot be obtained.

Let us start with the simplest example of measurement, where just one scalar variable is measured: \( x \), a stochastic variable with its own PDF \( P \). In order to account for variability in the computed moments, we postulate that \( P \) is a function, besides \( x \), of some parameters \( a_i \): \( P = P(x; [a_i]) \); hence \( \langle x^n \rangle = \int P x^n \, dx \) are functions of \( a_i \). The role of \( a_i \) is modelling the interaction between the system and its environment. Let us consider first the case when we have one parameter available: \( S = S(a) \), \( K = K(a) \) (the zero-parameter case is trivial, since no variability of \( S \) and \( K \) is then allowed). We make the rather natural postulate that the dependence on \( a \) is smooth. We may suppose that there exists a neighbourhood around \( S = 0 \) where the relation \( S(a) \) is reversible: \( a = a(S) \). Hence, \( K = K(a(S)) \rightarrow K(S) \). Because of the postulated smooth dependence on \( a \), we may Taylor expand \( K \) around \( S = 0 \): \( K = K_0 + K'S + (K''/2)S^2 + \cdots \). Many systems are invariant with respect to the sign inversion of \( x, x \rightarrow -x \). For instance, \( x \) may stand for the measurement of a velocity, as is often the case in fluid dynamics or geosciences. An inversion of sign corresponds to the arbitrary choice of the direction of motion. However, \( S \) is odd with respect to this operation, whereas \( K \) is even. Therefore, all coefficients in front of odd powers of \( S \) must be null in the above expansion. This yields

\[
K = K_0 + (K''/2)S^2
\]  

neglecting terms of order \( S^4 \) or higher. The presence of a linear contribution in \( (2) \) could be related to the breaking of the symmetry \( x \rightarrow -x \); that is, if some constraints do exist in the system preventing \( x \rightarrow -x \) from being an operation physically realizable for the system under consideration. For instance, if \( x \) stands for the measurement of a particle density, only positive values make sense. In this case, there does not appear to be any justification for discarding the linear term (but for the case of very small fluctuations, as we shall emphasize later).

Being a truncated Taylor expansion, equation (2) may only be valid as long as higher order terms remain negligible. Assessing how large is this range \( \Omega \) is an issue that, in principle, can be resolved only by actual inspection of each specific system. We provide, however, several considerations supporting the view that, for most systems, \( \Omega \) is as large as the region that may be experimentally scanned.

It is well known that, basing upon purely mathematical manipulations, one can provide a lower bound to \( K \), whatever the PDF [10]:

\[
K \geq S^2 + 1.
\]  

Equation (5) assigns upper and lower bounds achievable by \( S \). We may set \( |l|, u \geq 1 \) without much loss of generality (it must be \( u - l > 1 \) by construction). Hence, (5) reduces to \( l < S < u \). On the other hand, \( l \) and \( u \) cannot be exceedingly large, since this would imply that most of the data lie in a very narrow interval of values (remember that they are normalized to the variance); this is not the signature of a turbulent system.

An example of the region allowed to be spanned by a system in the plane \((S, K)\) is shown in figure 1.

For systems endowed with mirror symmetry (hence, \( -l = u \)), we may write the formal expression: \( K = G(S^2) + C(S^2) \) and, as \( S \rightarrow l, u \), because of (4), \( G(S^2) \rightarrow 1, C(S^2) \rightarrow 1 \). Empirically, \( G(0) = 3/2 \) [1, 6] and \( G(0) = 1.66 \) for the data in figure [2]; the estimate of \( C(0) \) is less precise due to the large vertical scatter of points, but is always of the order of unity. This conforms us in guessing that \( G \) and \( C \) remain almost constant throughout all the range \((l, u)\) and hence the parabolic relation between \( K \) and \( S \) must hold in the same interval.

As a second supporting piece of evidence, we note that disparate problems may be collapsed quite often to within a few classes of systems, amenable to analytical treatment. Accordingly, their empirical PDFs can be well approximated by a few classes of analytical functions. For example, lognormal PDF, Gamma PDF, etc arise in all those problems that involve addition of random variables \( X \) regardless of the nature of \( X \) (say, queuing models, the flow of items
through manufacturing and distribution processes and the load on web servers); the Poisson distribution is related to the waiting times between independent events, and so on. In the case of these analytical PDFs, the parameters \( a_i \) are simply the free parameters entering the definition of the PDF. We restrict to those analytical PDFs where \( S \) and \( K \) depend upon one single parameter. A scrutiny among these classes of analytical functions shows that equation (2) turns out to be an excellent approximation over at least a large \( S \) interval, and it is even an exact relation, valid for all \( S \). We mention, e.g., the problems that lead to the Gamma, inverse Gaussian, Poisson, \( \chi^2 \) and generalized extreme value (GEV) distributions. It is straightforward to check that the quadratic relation between \( S \) and \( K \) is exact when the dependence on the parameter \( a \) takes a fairly simple expression: \( S \) and \( K \) may be parameterized as \( S \propto a, K \propto a^2 \) plus constant terms. Therefore, although we reached (2) through a Taylor expansion, its validity is based on more general considerations: all those systems for which (I) the interaction with the environment may be modelled by means of just one effective continuously varying parameter \( a \), and (II) \( a \) can be defined such that the ‘response’ of the system, quantified in terms of departure of \( S \) from its Gaussian value, is linear with respect to \( a \), are expected to obey equation (2).

Let us move to the case where more parameters drive \( S \) and \( K \). We will study the two-parameter case. The other cases are not necessarily a trivial generalization of this one, but we believe that, even limiting to two parameters, we are able to include a large range of realistic systems. Let \( S = S(a_1, a_2), K = K(a_1, a_2) \). No such inversion as in the one-parameter case is possible now. However, we can still reduce to that case when the system is ‘quasi one-dimensional’, i.e. when the dependence on one of the parameters (say, \( a_1 \)) is much fainter that from the other \( (a_2) \). Hence, expressing \( a_2 = a_2(S(a_1), K = K(a_1, a_2(S(a_1))) \) leads to an expression formally identical to (2): \( K \approx K_0(a_1) + K'(a_1)S + (1/2)K''(a_1)S^2 + \ldots \). Therefore, for each fixed value of \( a_1 \), the curve \( K(S) \) is approximately parabolic. Varying \( a_1 \), we plot on the plane \((S, K)\) different parabolas. If the dependence on \( a_1 \) is strong, a scan over its admissible range of values will lead to plotting parabolas that span all or most of the plane. Conversely, if \( K_0, K' \) and \( K'' \) do depend only weakly upon \( a_1 \), we recover a fan of parabolas close to each other, practically spanning a restricted region of the \((S, K)\) plane. Therefore, the presence of a weak dependence from a second parameter may account for the spread of the points around the fitting parabola that is commonly observed in experiments (see, e.g., [1]). Actually, this spreading cannot be attributed to ‘experimental errors’ or other sources of noise: the statistical error due to the finiteness of the sample can be computed and is negligible in our cases. It is important therefore that our theory be able to explicitly take the spread into account.

We substantiate the above statements with a few examples where our conjectures may be verified explicitly. The first example involves the Hasegawa–Mima–Charney (HMC) equation, which in its non-dissipative version takes the form

\[
(1 - \nabla^2) \frac{\partial \phi}{\partial t} + \nabla \cdot \left( \frac{\partial \phi}{\partial y} \right) = -[\phi, \nabla^2 \phi] = 0,
\]

where \([\cdot, \cdot] \) stands for the Poisson bracket. HMC is a two-dimensional nonlinear equation widely known and studied for its capability of model wave behaviour of such widely different systems as electrostatic drift waves in magnetized plasmas and the incompressible motion of shallow rotating neutral fluids (see [12], chapter 6): \( \phi \) is the electrostatic potential in the first case and the fluctuation of the fluid depth in the second one. It is therefore a fairly good workhorse for a statistical theory of turbulence.

Horton and Ichikawa made a thorough analysis of this equation. Among its solutions (although not exhausting their whole range) they found that two scenarios may coexist. At low amplitudes of the field \( \phi \), fluctuations do resolve into a sea of linear non-interacting waves. The field \( \phi \) is therefore (even wildly) fluctuating but no longer technically turbulent. The central limit theory applies in this case: the measured signal is given by the sum of a large number of independent contributions and therefore has Gaussian statistical properties. The nonlinear high-amplitude part of the fluctuations develops into solitary coherent structures (vortices). In order to describe realistic systems interacting with an external environment, the dynamics must be augmented with source and dissipative terms, whose relative equilibrium will determine the average amplitude of fluctuations. Ultimately, therefore, we expect a whole continuum of its statistical properties, at one extremum including the Gaussian limit. The peculiar form of these solutions of the HMC equation allows for an explicit computation of its statistical moments ([12], par. 6.7). It is convenient to introduce the parameters \( f_p = \) packing fraction, that is, the fraction of surface occupied by nonlinear coherent structures; and \( A = \) amplitude of the nonlinear part of the fluctuations, which we suppose to be constant in order to grasp simpler results, and normalize with respect to linear fluctuations: \( \langle \phi^2 \rangle \equiv 1 \).

Finally,

\[
S \approx \frac{-3f_pA + 6f_pA^2}{(f_pA^2 + 1)^{3/2}}, \quad K \approx \frac{f_pA^4 + 6f_pA^2 + 3}{(f_pA^2 + 1)^2}.
\]

For fixed \( A \) and small \( f_p \) (say, \(< 0.1 \)), equation (7) yields a linear dependence between \( S \) and \( K \). The possibility of a linear dependence had to be envisaged because of the lack of symmetry \( A \rightarrow -A, f_p \rightarrow -f_p \). Conversely, the trend is almost quadratic with \( A \) for fixed \( f_p \) and moderately large \( A \) (< 5). It is interesting that, according to real data [13], \( f_p \) is actually small (<0.1–0.2). In realistic situations, several fields are coupled. For instance, in plasma physics the HM equation goes into the two-equation Hasegawa–Wakatani model when non-adiabatic small density fluctuations are included, too. Increasing the number of fields increases the number of control parameters, too, but each field \( F \) depends strongly only upon a subset \( a_F \) of all parameters, the remaining ones playing a weaker role; hence, qualitatively things are not different from the ‘quasi-one-dimensional’ case studied earlier.

The model proposed by SS [6] has played a major role in our considerations; hence it is interesting to see how it fits into our picture. We refer the reader to the original paper for the details and provide here just the fundamental results. The SS model reduces basically to just one equation for the time evolution of sea surface temperature fluctuations \( \delta T \) in the presence of external fluctuating forcing (heat flux). Part of the forcing \( (R) \) is due to background unknown
sources, and modelled as an additive noise. Another part \((F)\) arises as a consequence of the coupling between the sea surface and the atmosphere. Since feedback from the sea over the atmosphere cannot be neglected, \(F\) is function of \(\delta T\). The random character of the drive converts the evolution equation into a stochastic differential equation (SDE) with additive and multiplicative noise. The stationary PDF of the temperature fluctuations and all moments are retrieved by solving the associated Fokker–Planck equation, which can be done analytically. Four parameters are needed: \(\sigma_F\) and \(\sigma_R\) are characteristic times that quantify the rate of relaxation of \(\delta T\); \(F\) towards steady states (in [6], the parameter \(\phi\) is used: \(\phi_F = \phi^2 \times \sigma_F^2\)), but two are used by assigning the mean value and the variance. Hence, we remain with two parameters and we expect equation (1) to hold and \(A, B\) to be a function of one further control parameter \(a_0\). The explicit calculations of SS yield the exact result, confirming our expectations

\[
K = \frac{3}{2} \left(1 - \frac{a_0}{2}\right) S^2 + 3 \left(1 - a_0\right), \quad a_0 = \frac{\phi_F}{2\phi_R - \lambda}.
\]

Summarizing, we claim that the parabolic relation (1) arises because of the validation of the following conditions: (A) fluctuating systems may ultimately be modelled by SDE; (B) the interaction with the external environment is phenomenologically fed into the SDE through a number of effective parameters that, ultimately, enter into the definition of the moments \(S, K\). It is often possible to identify a single parameter \(S\), that \(K\) depends strongly on, and a small number of secondary parameters. This does not appear to be an exceedingly demanding requisite. In contrast, most if not all systems are modelled through equations that depend on a very small number of parameters. The paradigm is the Navier–Stokes equation that, once in dimensionless form, admits as external parameter the Reynolds \(Re\) number. We were not able to find an explicit study of skewness versus kurtosis for fluid turbulence driven by the Navier–Stokes equation, but several researchers addressed the issue of the scalings \(S, K\) versus \(Re\). For a review with data, see [7]. Figures 5 and 6 of that paper show that both \(S\) and \(K\) scale with \(Re\) for rather large values of this parameter (\(Re > 100\)): \(S, K \propto Re^{1.85a_2}a_2\). Hence, \(K \propto S^{1.85a_2}\), and \(a_2\) is rather close to 2 (the best fit being 2.5).

(C) Under a very weak external drive, many (although definitely not all) systems collapse to their nonturbulent Gaussian limit, a linear superposition of independent oscillations, and correspondingly \(S = 0, K = 3\). Finite driving smoothly displaces the system from this condition, and the resulting signal is a combination of Gaussian and nonlinear non-Gaussian fluctuations. The driving itself must be implemented into a parameter, \(a\). A linear response ansatz makes \(S\) proportional to \(a\). This, together with the next point (D), leads to the exact validity of (2). (D) The linear term in (1) is absent because the system studied has intrinsic mirror symmetry \(x \rightarrow -x\). Another possibility is that only small fluctuations \(\delta x\) around an equilibrium state are investigated. In this latter case, the full system may not possess mirror symmetry, but the reduced one does: \(\delta x \rightarrow -\delta x\). In this case, one may be led to condition (C) if the zero-fluctuation limit of the system is not turbulent and the small fluctuations make the system slightly depart from Gaussian statistics, making equation (2) sensible. (E) Finally, purely mathematical constraints exist, arising just out of the definition of \(S, K\) and from the fact that physically realizable systems are always finite, which prevents in any case this couple of parameters departing sensitively from the scaling (1).

In summary, therefore, our claim is that the parabolic relation between \(S\) and \(K\) encountered in the statistical treatment of data from turbulent systems is not likely to provide relevant information about the underlying physics.

Acknowledgments

This work was supported by the European Communities under the contract of the Association between EURATOM and ENEA. D Escande and E Martines read several versions of the draft and suggested invaluable corrections and improvements.

References

[1] Labit B et al 2007 Phys. Rev. Lett. 98 255002
[2] Agostini M 2009 Study of edge turbulence in magnetically confined fusion plasmas PhD Thesis unpublished
[3] Maurizi A 2006 Nonlinear Process. Geophys. 13 119
[4] Alberghi S, Maurizi A and Tampieri F 2002 J. Appl. Meteorol. 41 885
[5] Schopflocher T P and Sullivan P J 2005 Bound.-Layer Meteorol. 115 341
[6] Sura P and Sardeshmukh P D 2008 J. Phys. Oceanogr. 38 638
[7] Sreenivasan K R and Antonia R A 1997 Ann. Rev. Fluid Mech. 29 435
[8] Bertelrud A, Johnson S, Lytle C and Mills C 1997 Proc. Int. Cong. on Instrumentation in Aerospace Simulation Facilities http://ieeexplore.ieee.org
[9] Krommes J A 2008 Phys. Plasmas 15 030703
[10] Pearson K 1916 Phil. Trans. R. Soc. A 216 432
[11] Simpson J A and Welch B L 1960 Biomektrika 47 399
[12] Teuscher F and Guirard V 1995 Stat. Probab. Lett. 22 257
[13] Horton W and Ichiowaki Y H 1996 Chaos and Structures in Nonlinear Plasmas (Singapore: World Scientific)
[14] Spolaore M et al 2004 Phys. Rev. Lett. 93 215003
[15] Agostini M et al 2007 Phys. Plasmas 14 102305