Annals of Physics 363 (2015) 164–184

Contents lists available at ScienceDirect

Annals of Physics

journal homepage: www.elsevier.com/locate/aop

How random is a random vector?

Iddo Eliazar

Smart Device Innovation Science Team, New Devices Group, Intel Corporation, Yakum, Israel

ARTICLE INFO

Article history:
Received 12 June 2015
Accepted 23 September 2015
Available online 1 October 2015

Keywords:
Wilks' generalized variance and standard deviation
Perplexity
Uncorrelation and independence indices
Randomness measures and randomness diagrams
Lévy laws
Random walks

ABSTRACT

Over 80 years ago Samuel Wilks proposed that the "generalized variance" of a random vector is the determinant of its covariance matrix. To date, the notion and use of the generalized variance is confined only to very specific niches in statistics. In this paper we establish that the "Wilks standard deviation" – the square root of the generalized variance – is indeed the standard deviation of a random vector. We further establish that the "uncorrelation index" – a derivative of the Wilks standard deviation – is a measure of the overall correlation between the components of a random vector. Both the Wilks standard deviation and the uncorrelation index are, respectively, special cases of two general notions that we introduce: "randomness measures" and "independence indices" of random vectors. In turn, these general notions give rise to "randomness diagrams"—tangible planar visualizations that answer the question: How random is a random vector? The notion of "independence indices" yields a novel measure of correlation for Lévy laws. In general, the concepts and results presented in this paper are applicable to any field of science and engineering with random-vectors empirical data.

1. Introduction

Our everyday world is a Newtonian one, yet it is far from being deterministic. Randomness is ubiquitous, and it is inherent in practically every quantity we can think of. Indeed, from minute variations of seemingly identical items produced by a state-of-the-art manufacturing line, to wild fluctuations in the prices of financial assets, randomness can be small or large, but it is effectively unavoidable and hence always present.

E-mail addresses: iddo.eliazar@intel.com, eliazar@post.tau.ac.il.

http://dx.doi.org/10.1016/j.aop.2015.09.023

The most pervasive quantitative measure of randomness is the standard deviation, which is defined in the context of real-valued quantities, i.e. real-valued random variables. However, quantities of interest are most often multi-dimensional rather than one-dimensional, thus yielding random vectors rather than random variables. This leads us to our first question: *What is the standard deviation of a random vector?*

The standard deviation of a random variable quantifies its fluctuations about its mean. This one-dimensional construction can be straightforwardly extended to a multi-dimensional setting — thus apparently answering the first question. Alas, the resulting extension of the standard deviation fails to meet properties that we would expect any sensible measure of randomness to satisfy. So, we are back to step one.

To answer the first question we take to the drawing board and devise a general notion of "*randomness measure*" in the context of random vectors. Gaining insight from entropy, as well as from the failed attempt to straightforwardly extend the one-dimensional standard deviation to higher dimensions, we arrive at the following result: the randomness measure that generalizes the standard deviation from random variables to random vectors is the *Wilks' standard deviation*.

In 1932 the American statistician Samuel Wilks proposed that the "generalized variance" of a random vector is the determinant of its covariance matrix [1–3]. The Wilks' generalized variance is applied in several specific niches in mathematical statistics (see [4]–[5] and references therein). Yet, although it was introduced over 80 years ago, the Wilks' generalized variance did not penetrate beyond these niches, and it is unknown to the overwhelming majority of scientists and researchers. The Wilks' standard deviation is the square root of the Wilks' generalized variance.

In the one-dimensional case the "randomness" of a random variable can very well be perceived as the magnitude of its fluctuations about its mean, as indeed quantified by the standard deviation. In the multi-dimensional case the "randomness" of a random vector can also be perceived as the degree of inter-dependences between the vector's components — being less random if its components are strongly correlated, and being more random if its components are weakly correlated.

In the context of two random variables, correlation is defined as the ratio of their covariance to the product of their standard deviations. Correlation is the most prevalent and widely applied measure of statistical dependence. However, quantities of interest are most often multi-dimensional rather than two-dimensional, thus yielding random vectors of arbitrary dimension. This leads us to our second question: *What is the correlation of a random vector?*

Beyond answering the first question, the Wilks' standard deviation turns out to further facilitate an answer to the second question. Indeed, we establish the following result: The "*uncorrelation index*" of a random vector is the ratio of its Wilks' standard deviation to the product of the standard deviations of its components. The "uncorrelation index" is a measure that takes values in the unit interval, yields its zero lower bound if and only if the random-vector's components are affinely dependent, and yields its unit upper bound if and only if the random-vector's components are uncorrelated.

As noted above, on route to the Wilks' standard deviation, and in the context of random vectors, we devised a general notion of "randomness measure". With a certain condition holding, a general randomness measure can be turned into a gauge of the inter-dependences of random-vectors' components: the "*independence index*" of a random vector is the ratio of its randomness level to the product of the randomness levels of its components. The "independence index" is a measure that takes values in the unit interval, yields its zero lower bound when the random-vector's components are affinely dependent, and yields its unit upper bound when the random-vector's components are statistically independent.

Altogether, this paper addresses the question "*how random is a random vector?*", presents a comprehensive study of this question, and yields practical and applicable answers. One goal of this paper is to popularize the use of the Wilks' standard deviation, and introduce it to wide scientific audiences. Another goal of this paper is to introduce the notions of "uncorrelation index" and "independence index", and push for their application in the analysis of empirical data. In particular, we propose using "randomness diagrams" to visually describe the randomness of different random vectors: a point on the x–y plane representing each random vector, with x being its Wilks' standard deviation (or general randomness measure), and with y being its uncorrelation index (or general independence index).