Annals of Physics 374 (2016) 138–161

Contents lists available at ScienceDirect

Annals of Physics

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

From micro-correlations to macro-correlations

Iddo Eliazar

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

ARTICLE INFO

Article history:
Received 9 January 2016
Accepted 19 July 2016
Available online 17 August 2016

Keywords:
Overall correlation
Micro correlation
Macro correlation
Mixture models

ABSTRACT

Random vectors with a symmetric correlation structure share a common value of pair-wise correlation between their different components. The symmetric correlation structure appears in a multitude of settings, e.g. mixture models. In a mixture model the components of the random vector are drawn independently from a general probability distribution that is determined by an underlying parameter, and the parameter itself is randomized. In this paper we study the *overall correlation* of high-dimensional random vectors with a symmetric correlation structure. Considering such a random vector, and terming its pair-wise correlation "micro-correlation", we use an asymptotic analysis to derive the random vector's "macro-correlation" : a score that takes values in the unit interval, and that quantifies the random vector's overall correlation. The method of obtaining macro-correlations from micro-correlations is then applied to a diverse collection of frameworks that demonstrate the method's wide applicability.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Assume that, given a high-dimensional random vector, you are asked to evaluate the *overall correlation* between the vector's components. In case of a two-dimensional random vector the answer is straightforward. Indeed, the vector's overall correlation is the correlation between its single pair of components. But what is the answer in the case of a random vector of arbitrary dimension d? And, moreover, what is the answer when the dimension d is very large?

Well, as in the case of two-dimensional random vectors, we can consider the pair-wise correlations between the different components of the given random vector. There are $d(d-1)/2$

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

http://dx.doi.org/10.1016/j.aop.2016.07.027
0003-4916/© 2016 Elsevier Inc. All rights reserved.

such correlations, and these correlations constitute the vector's correlation matrix. But how can we evaluate the vector's overall correlation from its correlation matrix? Alternatively, given two random vectors, how can we determine which is more correlated than the other?

Given two bodies, we can measure the temperature of each, and determine which is hotter and which is colder. Analogously, we are looking for a one-dimensional measure of overall correlation in the context of random vectors. Such a measure exists, it emanates from the notion of *Wilks' generalized variance* [1–5], and it is based on the determinant of the random vector's correlation matrix [6]. In particular, in the case of a two-dimensional random vector the overall correlation is the square of the correlation between the vector's pair of components.

As in the *Central Limit Theorem* (CLT) [7–9], and as in *Extreme Value Theory* (EVT) [10–12], our key focus in this paper is the evaluation of overall correlation in the high-dimensional limit $d \to \infty$. The CLT and EVT share a common bedrock model that considers the components of the given random vector to be independent and identically distributed (IID) random variables. The CLT addresses the aggregate of the vector's components, whereas EVT addresses the maximum of the vector's components. Both the CLT and EVT explore how the *microscopic* affects the *macroscopic*: the "microscopic" being the statistical distribution of the IID components; the "macroscopic" being the statistical distribution of the components' aggregate and maximum, in the high-dimensional limit $d \to \infty$.

Our goal in this paper – similarly to the CLT and EVT, yet in the context of correlation – is to explore how the microscopic affects the macroscopic. To that end the first step is to set a bedrock model. Evidently, the CLT and the EVT bedrock model of IID components is trivial and uninteresting, as it bears no dependencies and hence no correlations. Nonetheless, *mixture models* offer a neat way of transforming IID components to dependent components [13–15]. Specifically, in a mixture model the components are IID random variables that are drawn from a general probability distribution that is determined by an underlying parameter, and the underlying parameter is randomized before drawing the random variables. This randomization induces dependencies, and consequently mixture models yield a *symmetric correlation structure*: the pair-wise correlations between the different components of the given random vector share a common value.

Our bedrock model in this paper is the *symmetric correlation structure*. At first sight this bedrock model may appear somewhat too simple and restricting. However, this bedrock model is actually surprisingly rich. Indeed, along the paper we shall provide a diverse collection of frameworks that demonstrate many applications of this bedrock model. Examples of the applications to be presented include: Maxwell–Boltzmann, Bose–Einstein, and Fermi–Dirac statistics; Lévy bridges and Lévy random probabilities; the aforementioned mixture models; regular and anomalous diffusion; regular and anomalous relaxation; power-laws; and conditional Lévy processes.

In what follows, given a high-dimensional random vector with a symmetric correlation structure, we consider the vector's common pair-wise correlation value as our input, and term it "*micro-correlation*". Then, using the aforementioned measure of overall correlation together with an asymptotic analysis, we derive in explicit form an output and term it "*macro-correlation*". This output is a score that takes values in the unit interval, and that quantifies the overall correlation of the random vector of interest, in the high-dimensional limit $d \to \infty$.

Thus, to summarize: in this paper we present a method for obtaining macro-correlations from micro-correlations, in the context of high-dimensional random vectors with a symmetric correlation structure. The method is established in Section 2. A wide range of the method's applications is presented in Section 3, followed by a short conclusion in Section 4. The derivations of various results and formulae stated in Sections 2 and 3 are detailed in Section 5.

2. The method

The goal of this paper is to study the overall correlation of large systems with a symmetric correlation structure. In this section we describe the measurement of *overall correlation* and the notion of a *symmetric correlation structure*, and then present an asymptotic analysis that meets our goal.

Download English Version:

https://daneshyari.com/en/article/1855942

Download Persian Version:

https://daneshyari.com/article/1855942

Daneshyari.com