MODELING OF MULTIVARIATE INTERACTIONS THROUGH THEIR MANIFESTATIONS AND LOW DIMENSIONAL MODEL BUILDING VIA THE CUMULANT GENERATING FUNCTION

JHAN RODRÍGUEZ AND ANDRÁS BÁRDOSSY

Abstract. Growing dimensionality of data calls for beyond-pairwise interactions quantification. Measures of multidimensional interactions quantification are hindered, among others, by two issues: 1. Interpretation difficulties, 2. the curse of dimensionality. We propose to deal with interactions via subject-matter specific interaction manifestations. Interactions manifestations do not necessarily help to build a model. Hence, we provide (still another) rationale to consider joint cumulants as interactions measures or parameters. We exhibit how joint cumulants are connected to some illustrative interaction manifestations. We then suggest that joint cumulants can work as building blocks for low dimensional model building that considers interactions of high order. The approach resembles that of probability inversion in the area of expert knowledge based risk assessment, where a discrimination is made between "elicitation" variables, familiar to the experts, and "target" (or model) variables, consisting of the more abstract parameters of a mathematical model.

Introduction

Technological innovation has lead to a world full of data of an increasingly growing dimension. These data in turn contain important information, the extraction of which is an important task of Statistics (c.f. Lindsay et al. (2004)). An important type of information is the kind of interdependence among variables being represented by data. This calls for statistical means of extracting, quantifying and, if possible, modeling such interdependence.

The most typical coefficient of dependence quantification is the product-moment correlation coefficient. Other important coefficients include Spearman's ρ, Kendall's τ, Ginni's γ, Blonqvist's q, etc. (The reader is referred to Joe (1989) for more coefficients).

Efforts have been advanced for considering the interaction of more than just two variables at a time. Concerning measures built around the concept of copulas, Wolff (1980) presents a measure of dependence which can be considered an extension to Spearman's ρ, whereas Schmid et al. (2010) extend this work and introduce a series of measures that can be considered as extension to other well-established, two dimensional measures of dependence.

Another course of action, traceable back to Linfoot (1957), is to use entropy or mutual information as association coefficient. Joe (1989) proposes a number of measures of this type that apply to more than two variables, Peña and Linde (2007) introduces a measure which adjusts itself to dimension, so as to compare the intensity of association of two vectors of different dimension. Micheas and Zografos (2006) deal with the general case of φ-dependence, of which mutual information

is one particular case. The intensity of association is measured in terms of the deviance of the joint distribution from the distribution given by the product of the marginal distributions (the independence case).

From the applied point of view, interactions among more than two variables are of interest in neuronal sciences (Staude et al. (2010a,c,b), and also Grün and Rotter (2010), chapter 12), in hydrological research (Bárdossy and Pegram (2009, 2012)) and in empirical finance (Dhaene and Linders (2012); Dhaene et al. (2013)). The need to consider methods to diagnose and quantify interactions among more than two variables simultaneously has been identified by these authors as a potential substantial improvement for current methods.

This paper deals with some of the issues inherent to high dimension interaction measures and proposes an approach for dealing with these problems. Section 1 introduces some issues that one encounters when dealing with measures of interaction for more than two variables. Section 2 states the approach we suggest for dealing with these issues: to discriminate between interaction "parameters" and interaction "manifestations". Section 3 introduces joint cumulants and Lancaster Interactions. The relation between the two is exhibited. Section 4 exhibits the relation between joint cumulants and some illustrative interaction manifestations, as defined in this paper. Section 5 illustrates the ideas presented, in that a specific model is presented. In section 6, discussion of the results and some work in progress is provided.

1. Difficulties of defining a measure of multivariate interaction

1.1. **Interpretability.** It is relatively easy to imagine one variable influencing another (say, hight of father on hight of son), and to visualize association between two variables (e.g. with the aid of a dispersion plot). By looking at several bi-variate datasets, and noticing the computed correlation coefficient, one can get a rough idea, at least, of what a correlation coefficient with a value of -0.8 stands for. The situation becomes more complicated when one has three or four variables at hand. Assuming we have some "correlation" measure among four variables: how is one supposed to interpret a value of $corr\,(X_1, X_2, X_3, X_4) = -0.8$? Can one visualize a dataset producing such a coefficient? Answers to these questions do not come to mind as easily as in the two-dimensional case.

It has been suggested that major advances in the science of statistics usually occur as a result of the theory-practice interaction (Box (1976)), and that the parameters of a model should have clear subject-matter interpretations (Cox et al. (1995)). However, the kind of parameters and quantities an applied scientist or engineer deals with, and which serve as criteria in his research is not likely to include such an abstract thing as $corr\,(X_1, X_2, X_3, X_4)$. Our point is that interaction parameters as mere abstract constructions will not find much application, unless one can "paraphrase" their meaning and relate it to the problem of study.

A similar issue has been noticed in the area of probabilistic risk assessment. When, due to the absence of data, expert knowledge must be included to assess the risk of a specific event, working mathematical models often have rather abstract parameters ("target variables") that can not be easily interpreted or paraphrased in terms of the physical quantities ("elicitation variables") the expert is acquainted with. This problem is exacerbated when the joint behavior of such parameters is to be assessed. Hence the expert is asked to express his uncertainty judgments in terms of *elicitation variables*, i.e. *observable quantities* within the area of his

expertise. A target variable set for the model is then recovered, such that the elicitation variables produced by the postulated model look as similar as possible like the elicited variables provided by the expert. This is an inverse problem, labeled "probabilistic inversion". The interested reader is referred for more details to (Bedford and Cooke (2001); Du et al. (2006)) and the references therein.

We suggest in section 2 a course of action that is analogous to "probabilistic inversion" for the problem of interactions quantification and modeling.

1.2. **High parametric dimensionality.** As dimension of the random vector under analysis increases, a naive use of interaction coefficients becomes prohibiting. For example, the correlation matrix of a 10-dimensional random vector is an array having 45 correlation coefficients. Assume symmetry on the variables with respect to the association coefficient (i.e. the order of the variables plays no role on the coefficient's value). If one wants to consider 3-wise, 4-wise and 5-wise "correlation coefficients", the corresponding arrays would have 450, 4500, and 45000 coefficients. Hence, it is necessary to be able to select judiciously the interaction parameters with which to work, and impose reasonable constraints on them.

One option would be to compute these multivariate "correlation coefficients" for relatively low-dimensional marginal distributions, say, sub-vectors of dimension 4, 5 or 6. It would then be necessary to construct a model for the whole random vector, which "glues" together these 4-6 dimensional marginals, and incorporates the estimates of the high-order "correlation" coefficients into an overarching model.

In the area of Spatial Statistics (see, for example Cressie (1991); Cressie and Wikle (2011); Diggle and Ribeiro (2007)), where the random vector $\mathbf{X}$ spans hundreds or thousands of components, such an approach is customary. In this case, each component of $\mathbf{X}$ is labeled with a spatial location. The dependence structure of the vector is built from dependence coefficients between pairs of variables: covariances, correlation coefficients or rank correlation coefficients. A model in which a covariance matrix defines the interdependence among the variables is then assumed, such as the Normal model. A simple Spatial Model considers the covariance between every two components of a vector as a function of the distance between their labeling locations. This function must ensure positive definiteness of the covariance matrix so built. There are a few common covariance functions mostly used in practice, one of them is the powered exponential covariance function,

$$Cov\left(d\right) = \sigma_0^2 . I\left(d = 0\right) + \sigma_1^2 \exp\left(-\left(d/\theta_1\right)^{\theta_2}\right) \tag{1.1}$$

where $\theta_1 > 0$, $0 < \theta_2 \leq 2$, $\sigma_0^2 \geq 0$, $\sigma_1^2 \geq 0$ are the covariance function parameters. This function defines the covariance between every two components, X_i and X_j, of $\mathbf{X}$ such that their labeling locations are at a distance $d > 0$ from each other. With adequate estimators of $\left(\theta_1, \theta_2, \sigma_0^2, \sigma_1^2\right)$, the covariance between every pair of components of the vector can be found. If the vector is to be extended to a new component representing a new location, one can readily extend the covariance matrix for this new component.

A similar approach, whereby one integrates together the model on the basis of 4-6 marginals distributions is recommended below.

Another aspect that can be considered a sort of "curse" of dimensionality, is the coefficient of interdependence to use: there are too many features that multivariate datasets can exhibit. In the one-dimensional case, parameters such as mean, standard deviation, skewness and kurtosis (basically, the first four cumulants) give

a lot of information about the distribution of data, provided these data come from an unimodal distribution. Those parameters (mean, skewness coefficient, etc.) describe data to some extent, since they can be readily connected to specific questions about data: the location of data, how informative this location about data is, how symmetric the distribution is, to what extend can one expect values very far away from the mean. As a reference one may have in mind these characteristics for the normal distribution. That is, these coefficients can be *interpreted*.

In the two dimensional case, as we have said, different types of correlation coefficients are still interpretable. Actually, applying the terminology of probabilistic inversion, in the Spatial Statistics model given above covariance coefficients between components are indeed the "elicitation variables" we are acquainted with, and intend to reproduce properly. The parameters of the covariance function take the place of "target variables".

As dimension increases, a single coefficient says less and less about the data features. Additionally, it becomes more and more difficult to use directly such "correlation" coefficients as "elicitation variables". This points to the need to select more intuition-appealing measures of interaction.

2. Interaction parameters versus interaction manifestations

The approach we suggest in this paper can be summarized as follows: first select an interaction "manifestation" relevant for the research in question. Then fit (low-dimensional) interactions "parameters" that make the fitted distribution reproduce, as close as possible, the observed interaction manifestation. This is entirely analogous to the elicitation/target variable discrimination put forth above, but here the interaction manifestations are to be estimated on the basis of available data.

By interaction manifestation, we mean any function of more than one component of the random vector analyzed, $\mathbf{X} \in \mathbb{R}^J$, which can be interpreted as relevant for the research objective. For the sake of illustration:

(1) The distribution of the sum of subsets of components of a random vector. In the context of financial analysis, this sum is readily interpreted as "risk".
(2) The joint distribution of subsets of components, or the probability of trespassing simultaneously a threshold defined for each component. This is useful in many applications. For example, in the context of series systems reliability, such trespassing probability is the probability of "failure".
(3) Differential entropy, any information-based dependence measure, or any of the copula-based generalizations to correlation measures studied by Schmid et al. (2010), of subsets of components. Depending on the specific research carried out, these may have subject-matter interpretations, or can readily provide the versed researcher of a specific area with a summary picture of the dependence in the data.

Interaction manifestations are interesting for the problem at hand, but they are not very helpful for building a model that integrates them, let alone a low-dimensional model. Suppose we had at hand interaction parameters or coefficients which:

(1) Provide us with an idea of the number of variables interacting within the random vector analyzed, $\mathbf{X} \in \mathbb{R}^J$.
(2) Can be somehow (functionally) connected with the interaction manifestations that are interesting for the research carried out.

(3) Can be built into a parametric or semi-parametric model. This would immediately open up the possibility of a low-dimensional model, via a judicious selection of assumptions and/or constraints on the interaction parameters.

Then we could proceed, in the manner of an inverse problem, as follows:

(1) We find data-based estimates or approximations to the interesting interaction manifestations
(2) We fit the interactions parameters so as to match best the observed interaction manifestations

In the next section, we introduce a reasonable interaction measure, and through it, a reasonable type of interaction parameter with which one can work along the lines above, namely the joint cumulant.

3. The Lancaster Interaction Measure and Joint Cumulants

3.1. **Lancaster Interactions.** We deal now with a function, called "additive interaction measure" or "Lancaster interaction measure", introduced by Lancaster (1969) and later modified by Streitberg (1990).

An additive interaction measure $\Delta F(\mathbf{X})$ is a signed measure determined by a given distribution $F(\mathbf{X})$ on $\mathbb{R}^J$. Its defining characteristic is that it is equal to zero for all $\mathbf{X} \in \mathbb{R}^J$, if $F(\mathbf{X})$ can be written as the non-trivial product of two or more of its (multivariate) marginal distributions (Streitberg (1990)). For example, if $J = 4$ and F can be written as $F_{124}F_3$, being F_{124} and F_3 the marginal distributions of (X_1, X_2, X_4) and X_3, respectively, then $\Delta F(\mathbf{X}) \equiv 0$, for all $\mathbf{X} \in \mathbb{R}^J$.

An alternative explanation is that $\Delta F \equiv 0$, if one subset of $\mathbf{X}$'s components is independent of another subset of components. If $\Delta F \equiv 0$, then F is said to be "decomposable".

Lancaster Interaction measure is defined by

$$\Delta F(\mathbf{X}) = \sum_{\pi} \left\{ \left((-1)^{|\pi|-1} (|\pi| - 1)! \right) F_{\pi}(\mathbf{X}) \right\} \tag{3.1}$$

where the sum is over all partitions, π, of index set $C = \{1, \ldots, J\}$.

An example will help clarify the notation: for index set $C = \{1, 2, 3, 4\}$ there are 15 partitions, three of which are: $\pi_1 = \{\{1\}, \{2\}, \{3, 4\}\}$, $\pi_2 = \{\{1, 4\}, \{2, 3\}\}$, $\pi_3 = \{\{1, 2, 3, 4\}\}$. Their cardinalities are $|\pi_1| = 3$, $|\pi_2| = 2$ and $|\pi_3| = 1$, respectively. In general, a set of J elements has a total of B_J possible partitions[1], where $B_0 = B_1 = 1$ and any subsequent $B_{k>1}$ can be found (see e.g. Rota (1964)) by the recurrence relation $B_{k+1} = \sum_{r=0}^{k} \binom{k}{r} B_r$. The reader is referred to the textbook of Aigner (2006) for more on partitions and their enumeration.

The symbol F_{π_1} is further to be interpreted as

$$F_{\pi_1}(\mathbf{X}) = F_1(X_1) F_2(X_2) F_{34}(X_3, X_4) \tag{3.2}$$

that is, the product of the (multivariate) marginal distributions defined by partition π_1. The same explanation holds at (3.1) for any of the B_J partitions, π, of index set $C = \{1, \ldots, J\}$.

It will be convenient to define partition operator J_π, to be applied to F for a given partition π, by

$$J_\pi F \to F_\pi \tag{3.3}$$

[1]The number B_J is often called Bell's number.

where F_π is as in the example at equation (3.2).

Streitberg (1990, 1999) shows an important result concerning ΔF: given a probability distribution function F, function ΔF as in (3.1) is the *only* function built as a linear combination of products of (multivariate) marginal distributions of F, such that $\Delta F(\mathbf{X}) := 0$, whenever one subset of $\mathbf{X}$'s components is independent of another components subset.

Since the interaction measure is defined in terms of a given distribution F, we can define the interaction operator:

$$\Delta = \sum_{\pi} \left\{ \left((-1)^{|\pi|-1} (|\pi| - 1)! \right) J_\pi \right\} \tag{3.4}$$

which, upon application to the distribution in question, returns the additive interaction measure.

3.2. **Joint Cumulants.** Moments and cumulants can be defined as constants summarizing important information about a probability distribution and sometimes, even determining it completely (cf. Kendall and Stuart (1969)). In this section we deal with random variables having a probability density function. The development is also valid for discreet distributions, under simple modifications. The reader is referred to Kendall and Stuart (1969); Muirhead (1982); Billingsley (1986); McCullagh (1987) for more details on moments and cumulants.

The Cumulant Generating Function (c.g.f.), $K_{\mathbf{X}}(\mathbf{t})$, of a random vector, $\mathbf{X} \in \mathbb{R}^J$, is defined as the logarithm of the moment generating function (m.g.f.),

$$K_{\mathbf{X}}(\mathbf{t}) = \log\left(M_{\mathbf{X}}(\mathbf{t})\right) = E\left(\exp\left(\sum_{j=1}^{J} t_j X_j\right)\right) \tag{3.5}$$

where $\mathbf{t} \in \mathbb{R}^J$, assuming these functions exist.

Joint cumulants are then defined to be the coefficients of the Taylor expansion for $K_{\mathbf{X}}(\mathbf{t})$,

$$K_{\mathbf{X}}(\mathbf{t}) \sim \sum_{r_1=0}^{\infty} \cdots \sum_{r_J=0}^{\infty} \frac{\kappa_{r_1,\ldots,r_J} . t_1^{r_1} \ldots t_J^{r_J}}{r_1! \ldots r_J!} \tag{3.6}$$

and hence can be found by differentiating $K_{\mathbf{X}}(\mathbf{t})$ and evaluating at $\mathbf{t} = \mathbf{0}$,

$$\kappa_{r_1,\ldots,r_J} = \frac{\partial^{r_1+\ldots+r_J}}{\partial^{r_J} t_J \ldots \partial^{r_1} t_1} K_{\mathbf{X}}(\mathbf{t}) \mid_{\mathbf{t}=\mathbf{0}} \tag{3.7}$$

where $r_j \geq 0$ is a non-negative integer. An important particular case is the covariance coefficient, or second order joint cumulant,

$$\frac{\partial^2}{\partial t_i \partial t_j} K_{\mathbf{X}}(t_i, t_j) \mid_{(t_i,t_j)=(0,0)} = cov(X_i, X_j)$$

The c.g.f. of a sub-vector $\mathbf{Y} = (X_{j_1}, \ldots, X_{j_k})$, with indexes in an index set, $j_i \in I$, can be readily found in terms of that of $\mathbf{X}$, by setting the indexes not corresponding to $\mathbf{Y}$ to zero:

$$K_{\mathbf{Y}}(\mathbf{s}) = \left(E\left(\exp\left(\sum_{i=1}^{k} s_i X_{j_i}\right)\right)\right) = \log\left(E\left(\exp\left(\sum_{j=1}^{J} g_j(\mathbf{s}) X_j\right)\right)\right) = K_{\mathbf{X}}(g(\mathbf{s}))$$

where $g : \mathbb{R}^k \to \mathbb{R}^J$, and

$$g_j(\mathbf{s}) = \begin{cases} 1, & j \in I \\ 0, & j \notin I \end{cases}$$

An alternative definition for joint cumulants uses product moments as departing point (see, for example, Brillinger (1974)). Let $\mathbf{X} \in \mathbb{R}^J$ be a random vector. For a set $(X_{j_1}, \ldots, X_{j_d})$ of $\mathbf{X}$´s components, where some sub-indexes j_r may be repeated, consider joint moments

$$E(X_{j_1} \ldots X_{j_d})$$

Consider partition operator J_π^*, analogous to (3.3), related to each partition π of $(j_1, \ldots, j_d)$. This operator converts $E(X_{j_1} \ldots X_{j_d})$ into the product of the factors determined by partition π.

For example, for $d = 4$, (j_1, j_2, j_3, j_3) and $\pi = \{\{1\}, \{2,3\}, \{4\}\}$, one has partition components $v_1 = \{1\}$, $v_2 = \{2,3\}$ and $v_3 = \{4\}$. Upon application of J_π^*, we have,

$$J_\pi^* E(X_{j_1} \ldots X_{j_4}) = E(X_{j_1}) E(X_{j_2} X_{j_3}) E(X_{j_3})$$

In the general case

$$J_\pi^* E(X_{j_1} \ldots X_{j_d}) = \prod_{v \in \pi} E\left(\prod_{j_r \in v} X_{j_r}\right)$$

The alternative definition of joint cumulants can now be given.

For random variables $(X_{j_1}, \ldots, X_{j_d})$, their joint cumulant of order d is given by,

$$cum(X_{j_1}, \ldots, X_{j_d}) := \sum_{\pi} \left\{ \left((-1)^{|\pi|-1} (|\pi| - 1)! \right) J_\pi^* \right\} E(X_{j_1} \ldots X_{j_d}) \tag{3.8}$$

Two examples are:

$$cum(X_1, X_2) = E(X_1 X_2) - E(X_1) E(X_2)$$

and

$$\begin{aligned} cum(X_1, X_2, X_3) = {} & E(X_1 X_2 X_3) - E(X_1 X_2) E(X_3) - E(X_1 X_3) E(X_2) \\ & - E(X_2 X_3) E(X_1) + 2E(X_1) E(X_2) E(X_3) \end{aligned}$$

Hence joint cumulants can be seen, from a mere formalistic point of view, to form a kind of higher order covariance coefficient. The second order joint cumulant is just the typical covariance coefficient.

3.3. **Relationship between Lancaster Interactions and Joint Cumulants.** The similarity between (3.1) and (3.8) might have already seemed "suspicious" to the reader. Indeed, if we concentrate for now on the case $\mathbf{X} \in \mathbb{R}^2$, then Lehmann (1966) reports that:

$$Cov(X_1, X_2) = cum(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F_{12}(x_1, x_2) - F_1(x_1) F_2(x_2)] \, dx_1 dx_2 \tag{3.9}$$

under the condition that $E\left(\left|X_1^{k_1} X_2^{k_2}\right|\right) < +\infty$, for $k_j = 0, 1$.

This equation is often called "Hoeffding's formula" since it was first discovered by Hoeffding (1940). Of course, the above equation can be written in terms of the Lancaster interaction measure (3.1), as

$$cum\left(X_1, X_2\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \Delta F\left(x_1, x_2\right) dx_1 dx_2 \tag{3.10}$$

It turns out that this equation can be extended to higher dimensions. Let $\mathbf{X} \in \mathbb{R}^J$ be a random vector. As shown by Block and Fang (1988), we have that (page 1808):

$$cum\left(\mathbf{X}\right) = (-1)^J \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \sum_{\pi} \left\{ \left((-1)^{|\pi|-1} \left(|\pi| - 1\right)! \right) F_{\pi} \right\} d\mathbf{X} \tag{3.11}$$

under the condition that $E\left(\left|X_j^J\right|\right) < +\infty$, for $j = 1, \dots, J$. Again, this is the same as saying that

$$cum\left(\mathbf{X}\right) = (-1)^J \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \Delta F\left(\mathbf{X}\right) d\mathbf{X} \tag{3.12}$$

Thus, joint cumulants are equal (up to a known constant) to the integral of Lancaster Interaction measure; they are "summary" or "integral" measures of additive interaction. To our knowledge, this connection had not been pointed out elsewhere.

It goes without much explanation that the joint cumulants of a random vector $\mathbf{X}$ vanish whenever a subset of the vector is independent of another, since then the integrating function is identically zero. This property is well-known and oftentimes the reason why joint cumulants are used in practice (e.g. in Brillinger (1974); Mendel (1991)). The inverse is true only if the distribution of $\mathbf{X}$ is determined by its moments, which may or may not be a reasonable assumption, depending on the application.

In particular, whenever we have $cum\left(X_{j_1}, \dots, X_{j_d}\right) \neq 0$, where no index j_k is repeated, this means that one cannot decompose the distribution of $\left(X_{j_1}, \dots, X_{j_d}\right)$; at least d variables within $\mathbf{X}$ are interacting simultaneously with each other.

Our contribution here is that joint cumulants are seen as the integral of the Lancaster interaction measure. Now as shown by Streitberg (1990), ΔF is the only additive measure, built very elementarily with the marginal distributions of the random vector, which vanishes whenever one subset of $\mathbf{X}$'s components is independent of another subset of components.

We have provided a theoretical basis for declaring joint cumulants "interaction parameters", and the cumulant generating function a "dependence structure". The functional character of the c.g.f. opens up the possibility of parametric modeling. It is just another way of defining a model, alternative to the density specification.

We shall see below, how the parameters of this model can be connected with some interesting interaction manifestations.

4. Interaction manifestations in terms of interaction parameters

4.1. **The Edgeworth Expansion and the Saddlepoint Approximation.** We recall first some well-known results about density approximation. Details for all topics of this sub-section can be found in Barndorff-Nielsen and Cox (1990); Kolassa

(2006); we present here just the approximations, in the context of a distribution having a probability density function.

The *Edgeworth Expansion* is a series expansion of the probability density and of the probability distribution in terms of the joint cumulants (performing as coefficients) and of the multivariate normal distribution (performing as basis function). We employ in this sub-section the shorthand notation for summations used in Barndorff-Nielsen and Cox (1990), in order to avoid an overflow of symbols in these pages. Arrays are represented by symbols with superscripts and under-scripts. For example a matrix is represented by $a^{i,j}$ or by b_{ij}. An array with three dimensions would be $c^{i,j,k}$ or d_{ijk}, and so on. The product of these symbols indicates summation along all dimensions for which the index is repeated. For example the term $\frac{1}{6\sqrt{n}}\kappa^{j_1,j_2,j_3}h_{j_1j_2j_3}$, to be used below, should be interpreted as

$$\frac{1}{6\sqrt{n}}\kappa^{j_1,j_2,j_3}h_{j_1j_2j_3} = \frac{1}{6\sqrt{n}}\sum_{j_1=1}^{J_1}\sum_{j_2=1}^{J_2}\sum_{j_3=1}^{J_3}\kappa^{j_1,j_2,j_3}h_{j_1j_2j_3} \tag{4.1}$$

where, for example,

$$\begin{aligned}\kappa^{j_1,j_2,j_3} &= cum\left(X_{j_1},X_{j_2},X_{j_3}\right)\\ \kappa^{j_1,j_2,j_3,j_4} &= cum\left(X_{j_1},X_{j_2},X_{j_3},X_{j_4}\right)\end{aligned}$$

We can now introduce the Edgeworth Expansion. Let $\mathbf{Z}\in\mathbb{R}^J$ be a random vector with probability density function f. Assume also, without loss of generality, that $\mathbf{Z}$ has mean a vector of zeros, a $J\times J$ covariance matrix $\kappa^{i,j}=\Gamma$, and joint cumulants $\left\{\kappa^{j_1,j_2,j_3}\right\},\left\{\kappa^{j_1,j_2,j_3,j_4}\right\},\ldots$. If we have a random sample of n i.i.d. random vectors with the same distribution as $\mathbf{Z}$, namely $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$, then we can form the average random vector $\mathbf{X}=\frac{1}{n}\sum_{i=1}^n\mathbf{Z}_i$. This latter random vector has a density function $f_{\mathbf{X}}$ which can be formally written as the following series expansion, in terms of the summation shorthand notation:

$$\begin{aligned}f_{\mathbf{X}}(\mathbf{x}) = \phi_\Gamma(\mathbf{x})\Big\{1 &+ \frac{1}{6\sqrt{n}}\kappa^{j_1,j_2,j_3}h_{j_1j_2j_3}(\mathbf{x};\Gamma) + \frac{1}{24n}\kappa^{j_1,j_2,j_3,j_4}h_{j_1j_2j_3j_4}\\ &+ \frac{1}{72n}\kappa^{j_1,j_2,j_3}\kappa^{j_4,j_5,j_6}h_{j_1j_2j_3j_4j_5j_6}(\mathbf{x};\Gamma)\Big\} + O\left(n^{-\frac{3}{2}}\right)\end{aligned} \tag{4.2}$$

Where ϕ_Γ is the multivariate Normal density function with zero mean and covariance matrix Γ, and $h_{j_1\ldots j_k}(\mathbf{x};\Gamma)$ represents the evaluation at $\mathbf{x}$ of the k-order Hermite polynomial determined by the identity

$$\phi_\Gamma(\mathbf{x})\,h_{j_1\ldots j_k}(\mathbf{x};\Gamma) = (-1)^k\frac{\partial^k\phi_\Gamma(\mathbf{x})}{\partial x_{j_1}\ldots\partial x_{j_k}} \tag{4.3}$$

Actually, $\phi_\Gamma(\mathbf{x})$ is a Normal approximation to $f_{\mathbf{X}}$, and the factors within brackets are often referred to as "correction terms".

It could be protested that we have considered only the case of an average $\mathbf{X}=\frac{1}{n}\sum_{i=1}^n\mathbf{Z}_i$ of random vectors. However, if the distribution of $\mathbf{Z}$ is unimodal and not wildly skewed or leptokurtic, then the Edgeworth Approximation given in 4.2 is often a good approximation in practice even with $n=1$, as we shall use it. After all, a random variable does not have to be the result of averaging n variables in order to have cumulants as such an average variable. This is the case of the chi-squared distribution with n degrees of freedom, for example, which can be interpreted as the sum of n standard Normal variables after raising each to the second power.

The usefulness of retaining the dependence on n is that we are reminded of when the Edgeworth Expansion is useful in practice: When the cumulants of $\mathbf{X}$, of which the density must be approximated, do not explode as their order increases, i.e. they behave as if $\mathbf{X}$ were approximately an average.

The Edgeworth expansion is practically accurate near the expected value of the distribution, but degenerates as one moves towards the tails of the distribution.

The *Saddlepoint Approximation*, also called "tilted" Edgeworth Approximation, is a more accurate approximation to the density of $\mathbf{X}$ at the tails, which we can apply if we know its cumulant generating function $K_{\mathbf{X}}(\mathbf{t})$. In the context of considering $\mathbf{X}$ as the mean of n copies of $\mathbf{Z}$, the relation between the cumulant generating functions is $K_{\mathbf{X}}(\mathbf{t}) = nK_{\mathbf{Z}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right)$. As mentioned above, we shall be using this approximations as if we were dealing with a variable being the average of $n = 1$ random variables. Thus we remove in the following the dependence on such an underlying n and work directly with $K_{\mathbf{X}}(\mathbf{t})$.

In order to to introduce the Saddlepoint Approximation, assume for a moment we are trying to find the Edgeworth Expansion not of $f_{\mathbf{X}}(\mathbf{x})$, but of a related family of density functions, defined in terms of an auxiliary vector $\lambda \in \mathbb{R}^J$,

$$f_{\mathbf{X}}(\mathbf{x};\lambda) = \exp\left(\mathbf{x}^{\mathbf{T}}.\lambda - K_{\mathbf{X}}(\lambda)\right) f_{\mathbf{X}}(\mathbf{x}) \tag{4.4}$$

The idea is, for *each* $\mathbf{x} \in \mathbb{R}^J$ to choose the most advantageous value $\hat{\lambda}$ of $\lambda \in \mathbb{R}^J$ in order to make the Edgeworth approximation $\hat{f}_{\mathbf{X}}(\mathbf{x};\lambda)$ to $f_{\mathbf{X}}(\mathbf{x};\lambda)$ as accurate as possible. Of course, this will provide automatically an approximation for $f_{\mathbf{X}}$,

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = \exp\left(K_{\mathbf{X}}\left(\hat{\lambda}\right) - \mathbf{x}^{\mathbf{T}}.\hat{\lambda}\right) \hat{f}_{\mathbf{X}}\left(\mathbf{x};\hat{\lambda}\right)$$

which is in fact what we want.

The optimum value $\hat{\lambda}$ can be proved to be the one fulfilling $\mathbf{x} = \nabla K_{\mathbf{X}}\left(\hat{\lambda}\right)$, for the particular $\mathbf{x} \in \mathbb{R}^J$ in question, because then density $f_{\mathbf{X}}\left(\mathbf{x};\hat{\lambda}\right)$ corresponds to a random vector having its mean at $\mathbf{x}$, where the Edgeworth Approximation is most accurate. Now, under suitable regularity conditions, the leading term of the Edgeworth expansion of $f_{\mathbf{X}}\left(\mathbf{x};\hat{\lambda}\right)$ is a multivariate Normal density with covariance matrix with entries

$$\left(\hat{\Sigma}_{i,j}\right) = \frac{\partial^2 K_{\mathbf{X}}(\lambda)}{\partial\lambda_i \partial\lambda_j} \mid_{\lambda=\hat{\lambda}}$$

evaluated at its mean; that is,

$$f_{\mathbf{X}}\left(\mathbf{x};\hat{\lambda}\right) \approx \frac{e^0}{(2\pi)^{J/2} \det(\Sigma)^{1/2}}$$

Thus, the looked for approximation is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \exp\left(K_{\mathbf{X}}\left(\hat{\lambda}\right) - \mathbf{x}^{\mathbf{T}}.\hat{\lambda}\right) f_{\mathbf{X}}\left(\mathbf{x};\hat{\lambda}\right) \approx \frac{\exp\left(K_{\mathbf{X}}\left(\hat{\lambda}\right) - \mathbf{x}^{\mathbf{T}}.\hat{\lambda}\right)}{(2\pi)^{J/2} \det\left(\hat{\Sigma}\right)^{1/2}} \tag{4.5}$$

The error of this approximation is of order $O\left(n^{-1}\right)$ for all $\mathbf{x} \in \mathbb{R}^J$, if the joint cumulants of random vector $\mathbf{X}$ behave like an average of n iid random vectors. Suitable normalization can bring this order down to $O\left(n^{-2}\right)$.

In spite of the apparent disadvantage of having to re-compute the density estimation for each $\mathbf{x}$, the computational cost becomes considerably smaller than that of the Edgeworth Approximation as dimension increases, since the number of multivariate Hermite polynomials at 4.2 to evaluate increases exponentially with the dimension of $\mathbf{x}$.

The fact that we can approximate the probability density function of $\mathbf{X}$ in terms of its joint cumulants or c.g.f. points out already that a wide spectrum of characteristics of $\mathbf{X}$ can be approximated on the basis of these mathematical objects.

4.2. **Connection of dependence structure with interaction manifestations.** We shall see now explicitly the connection of joint cumulants and the c.g.f. with three of the interaction manifestations listed at section 2, which manifestations refer to subsets of components, $(X_{j_1},\ldots,X_{j_k})$, $1 \leq k \leq J$, of the random vector $\mathbf{X} \in \mathbb{R}^J$. Namely: the distribution of the sum of components; parameters related to the joint probability of the components; and the differential entropy of the components.

A relevant point here is that, except for the distribution of the sum of components, even with a lot of data at hand, estimation of the interaction manifestations mentioned can be done only for (multivariate) marginals of relatively low dimension, such as k equal to 3, 4 or 5. But if we had somehow a sensible c.g.f. at hand, these manifestations can be consistently integrated into the whole distribution (in much the same way as covariances are integrated into a Spatial Statistics model that spans thousands of variables) with the aid of the overarching dependence structure, that is, the c.g.f.

Assume for the moment you have a reasonable type of c.g.f., that is, one that seems reasonable for the problem at hand (for an illustration see section 5).

4.2.1. *Connection of dependence structure with Sums of components.* Given a random vector $\mathbf{X} \in \mathbb{R}^J$ representing the variables under analysis, we are interested in the distribution of variable $S_{\mathbf{X}} = \sum_{i=1}^{k} X_{j_i}$, where $(X_{j_1},\ldots,X_{j_k})$, $1 \leq k \leq J$, is a sub-vector of the random vector $\mathbf{X} \in \mathbb{R}^J$.

One approach is to find the cumulants of $S_{\mathbf{X}}$ in terms of the joint cumulants of $\mathbf{X}$, and then approximate the density of $S_{\mathbf{X}}$ using the Edgeworth Expansion. Since $S_{\mathbf{X}}$ is a one-dimensional random variable, one can alternatively find research-relevant quantiles of its distribution by inverting the Edgeworth Expansion, i.e. by using the Cornish-Fisher inversion.

To find the cumulants of $S_{\mathbf{X}}$, note that two of the properties of joint cumulants are Brillinger (1974): symmetry and multi-linearity. Symmetry means that $cum\left(X_{j_1},\ldots,X_{j_k}\right) = cum\left(P\left(X_{j_1},\ldots,X_{j_k}\right)\right)$ for any permutation $P\left(j_1,\ldots,j_k\right)$ of the indexes $(j_1,\ldots,j_k)$. Concerning multi-linearity, for any random variable $Z \in \mathbb{R}$, one has

$$cum\left(Z + X_{j_1},\ldots,X_{j_k}\right) = cum\left(Z,\ldots,X_{j_k}\right) + cum\left(X_{j_1},\ldots,X_{j_k}\right)$$

Combining these two properties, it can be shown that

(4.6)

$$\kappa_r\left(S_{\mathbf{X}}\right) = cum\left(\underbrace{S_{\mathbf{X}},\ldots,S_{\mathbf{X}}}_{r}\right) = \sum_{i_1=1}^{k}\left[\sum_{i_2=1}^{k}\cdots\left[\sum_{i_r=1}^{k} cum\left(X_{j_{i_1}},\ldots,X_{j_{i_r}}\right)\right]\right]$$

where $\kappa_r(S_{\mathbf{X}})$ denotes the r-th cumulant of random variable $S_{\mathbf{X}} = \sum_{i=1}^{k} X_{j_i}$. Then the interesting quantiles of $S_{\mathbf{X}}$ can be (approximately) written in terms of the κ_r via the Cornish-Fisher inversion.

As the dimension k of the sub-vector increases, this approach becomes impractical, since the sum at (4.6) comprises too many elements. Fortunately, knowing the c.g.f. of $\mathbf{X}$ tells much about the c.g.f. of sums of its components.

In a somewhat more general context as before, given a random vector $\mathbf{X} = (X_1, \ldots, X_J)$, one can study the joint distribution of aggregated variables of the form:

$$\begin{aligned} \xi_1 &= \sum_{j_1 \in I_1} X_{j_1} \\ \xi_2 &= \sum_{j_2 \in I_2} X_{j_2} \qquad (4.7) \\ &\vdots \quad \vdots \quad \vdots \\ \xi_l &= \sum_{j_l \in I_l} X_{j_l} \qquad (4.8) \end{aligned}$$

where I_k, for $k = 1, \ldots, l$ represent non-overlapping index sets such that

$$I_1 \cup \ldots \cup I_l = \{1, \ldots, J\}$$

The cumulant generating function of the l-dimensional vector so obtained is given by

$$\begin{aligned} (4.9) \quad K_{\xi}(\mathbf{t}) &= \log\left(E\left(\exp\left(\mathbf{t}.\xi'\right)\right)\right) = \\ &\log\left(E\left(\exp\left(t_1\xi_1 + \ldots + t_l\xi_l\right)\right)\right) = \\ &\log\left(E\left(\exp\left(t_1\sum_{I_1} X_{j_1} + \ldots + t_l\sum_{I_l} X_{j_l}\right)\right)\right) = \\ &\log\left(E\left(\exp\left(g_1(\mathbf{t})X_1 + \ldots + g_J(\mathbf{t})X_J\right)\right)\right) = \\ &\log\left(E\left(\exp\left(g(\mathbf{t}).\mathbf{X}'\right)\right)\right) = K_{\mathbf{X}}(g(\mathbf{t})) \end{aligned}$$

Function $g : \mathbb{R}^l \to \mathbb{R}^J$ is a vector function defined by

$$\begin{aligned} g(\mathbf{t}) &= \left(g_1(\mathbf{t}), \ldots, g_J(\mathbf{t})\right) \\ g_j(\mathbf{t}) &= \mathbf{t}.\left(\mathbf{1}(j \in I_1), \ldots, \mathbf{1}(j \in I_l)\right)' \qquad (4.10) \end{aligned}$$

where

$$\mathbf{1}(j \in I_k) = \begin{cases} 1, & j \in I_k \\ 0, & j \notin I_k \end{cases}$$

It is hence possible to find the cumulant generating function of random vector $\xi \in \mathbb{R}^l$ in terms of that of the original vector $\mathbf{X} \in \mathbb{R}^J$. Then, if we know the c.g.f. of the original random vector $\mathbf{X}$, the cumulants, the cumulant generating function (and hence the approximate joint density of the aggregated variables, via Saddlepoint approximation at (4.5)) of $\xi \in \mathbb{R}^l$ can be found. In this way it is possible to deal with interaction manifestations of these aggregate variables, as

well. If we consider $l = 1$, then we see that we can attack successfully the problem posed by equation (4.6).

4.2.2. *Joint probabilities of (multivariate) marginals.* Given a sub-vector $\mathbf{Y} := (X_{j_1}, \ldots, X_{j_k})$ of $\mathbf{X}$, in order to find probabilities of the form

$$\Pr\left(X_{j_1} \geq x_{j_1}, \ldots, X_{j_k} \geq x_{j_k}\right)$$

one should in principle integrate expression (4.5), for the c.f.g. of $\mathbf{Y}$.

In the uni-variate case, it is a well-established practice Huzurbazar (1999) to employ instead an accurate approximation to that integral, which is due to Lugannani and Rice (1980). Namely, in the univariate case, we have:

$$F_X(x_0) \approx \int_{-\infty}^{x_0} \frac{\exp\left(K_X\left(\hat{\lambda}(x)\right) - x\hat{\lambda}(x)\right)}{(2\pi)^{1/2}\left(\frac{d^2K_X(\lambda)}{d\lambda^2} \mid_{\lambda=\hat{\lambda}(x)}\right)^{1/2}} dx \approx \Phi(r) + \phi(r)\left\{\frac{1}{r} - \frac{1}{q}\right\} \tag{4.11}$$

Where $\hat{\tau}$ is such that $K_X^{'}(\hat{\tau}) = x_0$, and:

$$\begin{aligned} r &= sign(\hat{\tau})\left\{2\left[\hat{\tau}x_0 - K_X(\hat{\tau})\right]\right\}^{\frac{1}{2}} \\ q &= \hat{\tau}\left\{\frac{d^2K_X(\lambda)}{d\lambda^2} \mid_{\lambda=\hat{\tau}}\right\}^{\frac{1}{2}} \end{aligned}$$

Thus, one must not perform the numerical integration at all.

For the multivariate case, Kolassa and Li (2010) have provided a generalization of the Lugannani-Rice formula, which produces an approximation to probability $\Pr(\mathbf{Y} \geq \mathbf{y})$ of order $O\left(n^{-1}\right)$, for $\mathbf{X} \in \mathbb{R}^J$. This formula is extremely complicated and writing it here will most likely obscure rather than clarify anything. Only the probability distribution function of a multivariate Normal distribution with covariance matrix given by

$$\Gamma_{ij} = \frac{\partial^2}{\partial t_i \partial t_j} K_{\mathbf{X}}(\mathbf{t}) \mid_{\mathbf{t}=\mathbf{0}}$$

must be computed. For this task there are accurate methods available for up to 20 dimensions Genz (1993).

If one intends to deal with vectors of dimension at most 5, corresponding to multidimensional marginals of the random field modeled, we consider more convenient to use numerical integration of (4.5). For higher dimensions it would be better to use the result of Kolassa and Li (2010) in order to avoid difficult and inaccurate integrations.

4.2.3. *Differential entropy.* Using the shorthand notation of 4.1, define $Z(\mathbf{x}) := \frac{1}{3!}\kappa^{j_1,j_2,j_3} h_{j_1 j_2 j_3}(\mathbf{x};\Gamma)$. Hulle (2005) studies an approximation to the differential entropy of $\mathbf{X}$, which utilizes only the first correction term in 4.2:

$$\begin{aligned}(4.12)\quad \int f_{\mathbf{X}}(\mathbf{x})\log\left(f_{\mathbf{X}}(\mathbf{x})\right)d\mathbf{x} &= H\left(\phi_{\Gamma}\right)-\int f_{\mathbf{X}}(\mathbf{x})\log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{\phi_{\Gamma}(\mathbf{x})}\right)d\mathbf{x}\\ &\approx H\left(\phi_{\Gamma}\right)-\int \phi_{\Gamma}(\mathbf{x})\left(1+Z(\mathbf{x})\right)\log\left(1+Z(\mathbf{x})\right)d\mathbf{x}\\ \approx H\left(\phi_{\Gamma}\right)-\int \phi_{\Gamma}(\mathbf{x})\left(Z(\mathbf{x})+\frac{1}{2}Z(\mathbf{x})^{2}\right)d\mathbf{x} &= H\left(\phi_{\Gamma}\right)-\frac{1}{12}\Big\{\sum_{j=1}^{J}\left(k^{j,j,j}\right)^{2}\\ &\quad +3\sum_{i,j=1,i\neq j}^{J}\left(\kappa^{i,i,j}\right)^{2}+\frac{1}{6}\sum_{i,j,k=1,i<j<k}^{J}\left(\kappa^{i,j,k}\right)^{2}\Big\}\end{aligned}$$

The value of $H\left(\phi_{\Gamma}\right)$ can be found in closed form, $H\left(\phi_{\Gamma}\right)=\frac{1}{2}\log\left(\det\left(\Gamma\right)\right)+\frac{J}{2}\log\left(2\pi\right)+\frac{J}{2}$. The approximation (4.12) is accurate to order $O\left(n^{-2}\right)$.

4.3. **Summarizing.** As we have seen in this section, joint cumulants provide us not only with a lower bound for the number of variables interacting within a vector; joint cumulants can also be connected with relevant interaction manifestations, that may have a specific subject-matter interpretation. The extent to which our model reproduces these interaction manifestations, can be used for model calibration or validation.

5. Illustration: Extending the Gaussian model

The multivariate Normal model is a widely applied model in multivariate analysis. A random vector $\mathbf{X}\in\mathbb{R}^{J}$ having mean vector $\mathbf{m}$ and covariance matrix Γ, has c.g.f. given by,

$$(5.1)\qquad K_{\mathbf{X}}(\mathbf{s})=\mathbf{s}.\mathbf{m}^{T}+\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^{T}$$

A similar c.g.f. was studied by Steyn (1993),

$$(5.2)\qquad K_{\mathbf{X}}(\mathbf{s})=\mathbf{s}.\mathbf{m}^{T}+\frac{c_{1}}{1!}\left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^{T}\right)+\frac{c_{2}}{2!}\left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^{T}\right)^{2}+\frac{c_{3}}{3!}\left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^{T}\right)^{3}+\ldots$$

Indeed, this c.g.f. reduces to that of the Gaussian model by setting $c_{1}=1$ and $c_{r>1}=0$. In order to avoid identifiability problems of the covariance matrix, we set $c_{1}=1$ and declare Γ to be a true covariance matrix. This model is treated in detail at Rodríguez and Bárdossy (2013), in the context of spatial statistics; it is shown at Rodríguez and Bárdossy (2013) how it covers a span of tail dependence going from zero (i.e. Normal) to that of the Student-t.

Joint cumulants of this random vector are found by differentiating 5.2, and evaluating at $\mathbf{s}=\mathbf{0}$. All joint cumulants of odd order, $cum\left(X_{j_{1}},\ldots,X_{j_{k}}\right)$, for$k$ odd, are zero. The non-zero joint cumulants are of the form:

$$
\begin{aligned}
cum\left(X_j\right) &= m_j \\
cum\left(X_{j_1}, X_{j_2}\right) &= \frac{c_1}{2}\left\{\Sigma_{j_1 j_2} + \Sigma_{j_2 j_1}\right\} \\
cum\left(X_{j_1}, \ldots, X_{j_4}\right) &= \frac{c_2}{2!}\frac{1}{2^2}\left\{\Sigma_{j_1 j_2}\Sigma_{j_3 j_4} + \Sigma_{j_1 j_3}\Sigma_{j_2 j_4} + \Sigma_{j_1 j_4}\Sigma_{j_2 j_3}\right\} \\
&\vdots \\
cum\left(X_{j_1}, \ldots, X_{j_{2r}}\right) &= \frac{c_{\frac{r}{2}}}{\frac{r}{2}!}\frac{1}{2^{\frac{r}{2}}}\left\{\sum_{j_1,\ldots,j_r=1}^{J} \Gamma_{j_1 j_2}\ldots\Gamma_{j_{r-1} j_r}\right\} \qquad (5.3)
\end{aligned}
$$

and so on. The indices of the covariances involved in each product appearing in the summation at (5.3) form a partition for the index set $\{j_1, \ldots, j_r\}$.

Each higher order joint cumulant of order $r > 2$ is (low-dimensionally) built on the basis of the covariance coefficients of the components involved and just an additional coefficient, c_r. This is the kind of judicious constraining/selection of joint cumulants advocated for in this paper.

Moreover, joint cumulants of increasing order can be fitted "orthogonally" by fitting the respective c_r coefficient, *without* modifying the lower dimensional joint cumulants. Hence mean and covariances can be fitted by finding estimates for $\mathbf{m}$, Γ and c_1, and additional features of the random vector $\mathbf{X}$ can be fitted by finding estimates for $c_2, c_3, c_4, \ldots$.

By "additional features" we mean relevant interaction manifestations. In this way, we enrich the Gaussian model, and make it more capable of reproducing important features of data. This points to the necessity of a two-step estimation procedure. In a first step, means and covariances are fitted. On a second step, parameters $c_{r>2}$ are fitted, such that the relevant interaction manifestation observed in data is best reproduced.

We deal in the example below with the interaction manifestations: 99% and 99.5% quantiles of the sum of components of $\mathbf{X}$.

5.1. **Illustration for interactions manifestations related to the sums of components.** Assume we have realizations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of $\mathbf{X} \in \mathbb{R}^J$, where $\mathbf{x}_i = (x_{i1}, \ldots x_{iJ})$. We choose for the illustration a sample size of $N = 10000$

We assume data is standardized, so that each marginal component has mean zero and variance 1. Then one can safely assume that $c_1 = 1$, $m_j = 0$, for $j = 1, \ldots, J$, and that Γ is a correlation matrix.

One must also decide at what order one wishes to truncate the expansion (5.2). For example, $c_{r>3} := 0$.

Step 1, correlation estimation. The components of Γ can be estimated by the sample correlation coefficients,

$$\hat{\Gamma}_{i,j} = \hat{\rho}_{i,j}$$

In this way, we obtain an estimate $\hat{\Gamma}$ for the correlation matrix.

Step 2, interaction manifestation fitting. If we are interested on the distribution of $S_{\mathbf{X}} = \sum_{j=1}^{J} X_j$, as would be the case in empirical finance applications, the c.g.f. of $S_{\mathbf{X}}$ is (see sub-section 4.2.1)

$$K_{S_{\mathbf{X}}}(t) = \left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^T\right) + \frac{c_2}{2!}\left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^T\right)^2 + \frac{c_3}{3!}\left(\frac{1}{2}\mathbf{s}\Gamma\mathbf{s}^T\right)^3 + \ldots \qquad (5.4)$$

where

$$\mathbf{s} = \underbrace{(t, \ldots, t)}_{J}$$

Compute the sample quantiles of the sum, $q_1 = \hat{q}_{90\%}(S_{\mathbf{X}})$ and $q_2 = \hat{q}_{99\%}(S_{\mathbf{X}})$. These comprise the interaction manifestations you would like to reproduce faithfully.

Now, using

$$K_{S_{\mathbf{X}}}(t) = \left(\frac{1}{2}\mathbf{s}\hat{\Gamma}\mathbf{s}^T\right) + \frac{c_2}{2!}\left(\frac{1}{2}\mathbf{s}\hat{\Gamma}\mathbf{s}^T\right)^2 + \frac{c_3}{3!}\left(\frac{1}{2}\mathbf{s}\hat{\Gamma}\mathbf{s}^T\right)^3 + \ldots \tag{5.5}$$

as our c.g.f., we can apply the Lugganani-Rice approximation (4.11) to form approximations to $F_{S_{\mathbf{X}}}$. These approximations, $\hat{F}_{S_{\mathbf{X}}}$, are a function of the unknown parameters c_2 and c_3. Note that these parameters do not affect the already fitted correlation matrix. Then, one can define the following objective function:

$$Z(c_2, c_3) = \left(\hat{F}_{S_{\mathbf{X}}}(q_1) - 0.9\right)^2 + \left(\hat{F}_{S_{\mathbf{X}}}(q_2) - 0.99\right)^2 \tag{5.6}$$

Upon minimization of Z, one has estimates of c_2 and c_3 which best reproduce the interesting interaction manifestation, in addition to reproducing correlations properly.

6. Discussion

An approach for considering interactions that go beyond correlations has been presented. We have seen that the discrimination between interactions "parameters" and interactions "manifestations" can help to circumvent two major problems one is confronted with, when attempting to quantify and model higher order interactions: the problem of interpretability, by working with subject-matter relevant manifestations of interdependence; and the problem of high dimensionality, by recoursing to joint cumulants as building blocks of a dependence model, which cumulants can be judiciously selected/constraint. By using the cumulant generating function, we are recoursing to a well-studied object: the characteristic function of a distribution.

As dimension of vector $\mathbf{X}$ increases, interactions of high order may be more and more difficult to assess. For example, a random vector having c.f.g. (5.2), with $c_1 = 1$, $c_r \approx 0$ for $2 \leq r \leq 3$ but then $c_{r \geq 4} \neq 0$, would have one and two dimensional marginals practically equal to those of a Guassian distribution. But the interaction coefficients of groups of 14 components or more will be much different, producing different interaction manifestations. The difference in the overall dependence structures may grow tremendously as the dimension of the random vector $\mathbf{X}$ grow (i.e. $J >> 2$), even though these fact may go totally unnoticed in the one and two dimensional marginal analysis of data.

In Rodríguez and Bárdossy (2013), these issues are dealt with and illustrated in the context of Spatial Statistics, where the issue of low dimensionality is essential, and where interaction manifestations can differ drastically between two models having very similar 1 and 2 dimensional marginals, due to the big dimension of the field.

Acknowledgments. This research forms part of the Ph.D thesis of the first author, which was funded by a scholarship of the German Academic Exchange Service (DAAD).

References

Aigner, M. (2006). *Diskrete Mathematik*. Vieweg+Teubner Verlag.

Bárdossy, A. and Pegram, G. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences Discussions*, 6(3):4485–4534.

Bárdossy, A. and Pegram, G. (2012). Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resources Research*, 48(9).

Barndorff-Nielsen, O. E. and Cox, D. R. (1990). *Asymptotic techniques for use in statistics*. Chapman and Hall, London [u.a.].

Bedford, T. and Cooke, R. (2001). *Probabilistic risk analysis: foundations and methods*. Cambridge University Press.

Billingsley, P. (1986). *Probability and measure*. Wiley series in probability and mathematical statistics. Wiley, New York, 2nd ed edition.

Block, H. W. and Fang, Z. (1988). A multivariate extension of hoeffding's lemma. *The Annals of Probability*, 16(4):1803–1820.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):pp. 791–799.

Brillinger, D. R. (1974). *Time series: data analysis and theory*. Holt, Rinehart, and Winston, New York.

Cox, D. R., Bayarri, M., Cuadras, C., Bernadro, J. M., Girón, F., Moreno, E., Keiding, N., Lindley, D., Pericchi, L., Piccinato, L., et al. (1995). The relation between theory and application in statistics. *Test*, 4(2):207–261.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Wiley.

Cressie, N. A. C. (1991). *Statistics for spatial data*. Wiley series in probability and mathematical statistics. Wiley, New York.

Dhaene, J. and Linders, D. (2012). The herd behavior index: A new measure for the implied degree of co-movement in stock markets. *Insurance: Mathematics and Economics*, 50(3):357 – 370.

Dhaene, J., Linders, D., Schoutens, W., and Vyncke, D. (2013). A multivariate dependence measure for aggregating risks. Open access publications from katholieke universiteit leuven, Katholieke Universiteit Leuven.

Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer.

Du, C., Kurowicka, D., and Cooke, R. (2006). Techniques for generic probabilistic inversion. *Computational Statistics and Data Analysis*, 50(5):1164 – 1187.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Sciences and Statistics*, 25:400–405.

Grün, S. and Rotter, S. (2010). *Analysis of parallel spike trains*. Springer, New York.

Hoeffding, W. (1940). Masstabinvariante korrelations-theorie. *Schriften Math. Inst. Univ. Berlin*, 5:181–233.

Hulle, M. M. V. (2005). Edgeworth approximation of multivariate differential entropy. *Neural Computation*, 17(9):1903–1910.

Huzurbazar, S. (1999). Practical saddlepoint approximations. *The American Statistician*, 53(3):pp. 225–232.

Joe, H. (1989). Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405):pp. 157–164.

Kendall, M. G. and Stuart, A. (1969). *The advanced theory of statistics Vol. 1, Distribution theory*. Griffin, London.

Kolassa, J. (2006). *Series approximation methods in statistics*, volume 88. Springer.

Kolassa, J. and Li, J. (2010). Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli*, 16(4):1191–1207.

Lancaster, H. O. (1969). *The chi-squared distribution*. Wiley, New York.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):1137–1153.

Lindsay, B. G., Kettenring, J., and Siegmund, D. O. (2004). A report on the future of statistics. *Statistical Science*, 19(3):pp. 387–407.

Linfoot, E. (1957). An informational measure of correlation. *Information and control*, 1(1):85–89.

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12(2):pp. 475–490.

McCullagh, P. (1987). *Tensor methods in statistics*. Chapman and Hall, London; New York.

Mendel, J. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE*, 79(3):278–305.

Micheas, A. C. and Zografos, K. (2006). Measuring stochastic dependence using φ-divergence. *Journal of Multivariate Analysis*, 97(3):765 – 784.

Muirhead, R. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons, New York.

Peña, D. and Linde, A. v. d. (2007). Dimensionless measures of variability and dependence for multivariate continuous distributions. *Communications in Statistics - Theory and Methods*, 36(10):1845–1854.

Rodríguez, J. and Bárdossy, A. (2013). Beyond correlation in spatial modeling. Technical report, Institute for Modelling Hydraulic and Environmental Systems, Universität Stuttgart.

Rota, G.-C. (1964). The number of partitions of a set. *The American Mathematical Monthly*, 71(5):498.

Schmid, F., Schmidt, R., Blumentritt, T., Gaißer, S., and Ruppert, M. (2010). Copula-based measures of multivariate association. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, volume 198, pages 209–236. Springer Berlin Heidelberg, Berlin, Heidelberg.

Staude, B., Grün, S., and Rotter, S. (2010a). Higher-order correlations and cumulants. In Grün, S. and Rotter, S., editors, *Analysis of Parallel Spike Trains*, pages 253–280. Springer US, Boston, MA.

Staude, B., Grün, S., and Rotter, S. (2010b). Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Frontiers in Computational Neuroscience*, 4.

Staude, B., Rotter, S., and Grün, S. (2010c). Cubic: cumulant based inference of higher-order correlations in massively parallel spike trains. *Journal of Computational Neuroscience*, 29(1-2):327–350.

Steyn, H. (1993). On the problem of more than one kurtosis parameter in multivariate analysis. *Journal of Multivariate Analysis*, 44(1):1 – 22.

Streitberg, B. (1990). Lancaster interactions revisited. *The Annals of Statistics*, 18(4):1878–1885.

Streitberg, B. (1999). Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models. *The Annals of Statistics*, 27(1):405–413.

Wolff, E. F. (1980). N-dimensional measures of dependence. *Stochastica: revista de matemática pura y aplicada*, 4(3):175–188.