



Cumulants of multiinformation density in the case of a multivariate normal distribution

Guillaume Marrelec*, Alain Giron

Laboratoire d'imagerie biomédicale (LIB), Sorbonne Université, CNRS, INSERM, F-75006, Paris, France
 Centre d'études en sciences des interactions (CESI) / Center for Interaction Science (CIS), F-75006, Paris, France



ARTICLE INFO

Article history:

Received 21 March 2019
 Received in revised form 5 August 2019
 Accepted 8 August 2019
 Available online 23 August 2019

Keywords:

Dependence
 Information
 Cumulant-generating function
 Cumulants
 Mutual information
 Multiinformation

ABSTRACT

We consider a generalization of information density to a partitioning into $N \geq 2$ subvectors. We calculate its cumulant-generating function and its cumulants in the particular case of a multivariate normal distribution, showing that these quantities are only a function of all the regression coefficients associated with the partitioning.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Let \mathbf{X} be a multivariate normal random variable with distribution $f(\mathbf{x})$, and $(\mathbf{X}_1, \mathbf{X}_2)$ a partitioning of \mathbf{X} into 2 subvectors with corresponding marginals $f_1(\mathbf{x}_1)$ and $f_2(\mathbf{x}_2)$. The information density relative to \mathbf{X} and the partitioning $(\mathbf{X}_1, \mathbf{X}_2)$ is the random variable defined as (Polyanskiy and Wu, 2017, §17.1)

$$i_d(\mathbf{X}; \mathbf{X}_1, \mathbf{X}_2) = \ln \frac{f(\mathbf{X}_1, \mathbf{X}_2)}{f_1(\mathbf{X}_1)f_2(\mathbf{X}_2)}. \tag{1}$$

One of the key features of information density is that its expectation yields mutual information (Kullback, 1968, Chap. 1, §2; Polyanskiy and Wu, 2017, §17.1). In the present paper, we consider a partitioning of \mathbf{X} into $N \geq 2$ subvectors $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ with corresponding marginals f_1, \dots, f_N and define *multiinformation density* as

$$i_d(\mathbf{X}; \mathbf{X}_1, \dots, \mathbf{X}_N) = \ln \left[\frac{f(\mathbf{X})}{\prod_{n=1}^N f_n(\mathbf{X}_n)} \right].$$

The mean of this quantity

$$I(\mathbf{X}; \mathbf{X}_1, \dots, \mathbf{X}_N) = \int f(\mathbf{x}) \ln \left[\frac{f(\mathbf{x})}{\prod_{n=1}^N f_n(\mathbf{x}_n)} \right] d\mathbf{x}$$

* Corresponding author at: Laboratoire d'imagerie biomédicale (LIB), Sorbonne Université – Campus des Cordeliers, Escalier A – 3e étage, 15, rue de l'École de Médecine, 75006 Paris, France.

E-mail address: guillaume.marrelec@inserm.fr (G. Marrelec).

is itself a generalization of mutual information known under different names: total correlation (Watanabe, 1960), multivariate constraint (Garner, 1962), δ (Joe, 1989), or multiinformation (Studený and Vejnarová, 1998). Our interest in i_d is driven by its close connection with mutual independence. Indeed, when the \mathbf{X}_n 's are mutually independent, multiinformation is classically equal to 0, but we also have $i_d \equiv 0$ and $\text{Var}(i_d) = 0$ (see online supplement, §1), yielding other statistical markers of independence. By contrast, dependence between the \mathbf{X}_n 's is a multivariate phenomenon that multiinformation, as a one-dimensional measure, can only partially quantify. We expect i_d to give a more detailed characterization of dependence, e.g., through its moments or cumulants.

We here focus on multivariate normal distributions. The family of multivariate normal distribution with given mean $\boldsymbol{\mu}$ can be parameterized by either a covariance matrix or a concentration/precision (i.e., inverse covariance) matrix. Either parameterization shows multivariate distributions according to a certain perspective and emphasizes different features (e.g., Markov properties for the concentration matrix). In this context, we wished to investigate the existence of a natural way of parameterizing dependencies, i.e., a parameter that would emphasize the dependence properties of the distribution.

The core of the present paper is the following theorem:

Theorem 1. *Let \mathbf{X} be a d -dimensional variable following a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Partition \mathbf{X} into N subvectors $(\mathbf{X}_1, \dots, \mathbf{X}_N)$, and set i_d the corresponding multiinformation density. Then the cumulant-generating function of i_d is given by*

$$\ln E(e^{t i_d}) = t I(\mathbf{X}_1; \dots; \mathbf{X}_N) - \frac{1}{2} \ln |\boldsymbol{\Gamma}_d - t \boldsymbol{\Gamma}|, \quad (2)$$

where

$$\boldsymbol{\Gamma} = \boldsymbol{\Sigma} \text{diag}(\boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{NN})^{-1} - \boldsymbol{\Gamma}_d$$

is the block matrix whose diagonal blocks are equal to $\mathbf{0}$ and where each off-diagonal block (m, n) is the matrix of regression coefficients of \mathbf{X}_m on \mathbf{x}_n

$$\boldsymbol{\Gamma}_{m|n} = \boldsymbol{\Sigma}_{mn} \boldsymbol{\Sigma}_{nn}^{-1}, \quad m \neq n. \quad (3)$$

The cumulants of i_d are given by

$$\kappa_1(i_d) = I(\mathbf{X}_1; \dots; \mathbf{X}_N) \quad \text{and} \quad \kappa_l(i_d) = \frac{(l-1)!}{2} \text{tr}(\boldsymbol{\Gamma}^l), \quad l \geq 2.$$

This theorem is proved in Section 2. In Section 3, we investigate some consequences of this result. Section 3 is devoted to the discussion.

2. Proof of theorem

2.1. Cumulant-generating function

We partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in accordance with the partitioning of \mathbf{X} , so that $\boldsymbol{\mu}_n$ is the expectation of \mathbf{X}_n and $\boldsymbol{\Sigma}_{mn}$ the matrix of covariances between \mathbf{X}_m and \mathbf{X}_n . Multiinformation between the \mathbf{X}_n 's yields

$$I(\mathbf{X}_1; \dots; \mathbf{X}_N) = \frac{1}{2} \ln \frac{\prod_{n=1}^N |\boldsymbol{\Sigma}_{nn}|}{|\boldsymbol{\Sigma}|}.$$

From there, we can express multiinformation density as

$$i_d = I(\mathbf{X}_1; \dots; \mathbf{X}_N) + j_d, \quad (4)$$

where j_d is defined as

$$j_d = \frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^t \boldsymbol{\Phi} (\mathbf{X} - \boldsymbol{\mu}). \quad (5)$$

and $\boldsymbol{\Phi}$ as

$$\boldsymbol{\Phi} = \text{diag}(\boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{NN})^{-1} - \boldsymbol{\Sigma}^{-1}.$$

Here, $\text{diag}(\boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{NN})$ stands for the block-diagonal matrix with diagonal blocks equal to the $\boldsymbol{\Sigma}_{nn}$'s. The moment-generating function of j_d yields

$$E(e^{t j_d}) = \int (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t (\boldsymbol{\Sigma}^{-1} - t \boldsymbol{\Phi})(\mathbf{x}-\boldsymbol{\mu})} d\mathbf{x}.$$

Since it can be shown that $\Sigma^{-1} - t\Phi$ is positive definite at least in a neighborhood of $t = 0$ (see online supplement, §2), the integrand is proportional to a multivariate normal distribution with mean μ and covariance matrix $\Sigma^{-1} - t\Phi$. Integration with respect to \mathbf{x} therefore yields

$$E(e^{j_d}) = \frac{|\Sigma^{-1}|^{\frac{1}{2}}}{|\Sigma^{-1} - t\Phi|^{\frac{1}{2}}} = |\mathbf{I}_d - t\Gamma|^{-\frac{1}{2}}$$

and

$$\ln E(e^{j_d}) = -\frac{1}{2} \ln |\mathbf{I}_d - t\Gamma|, \tag{6}$$

where \mathbf{I}_d is the d -by- d unit matrix and $\Gamma = \Sigma\Phi$ the block matrix whose diagonal blocks are equal to $\mathbf{0}$ and where each nondiagonal block (m, n) is the matrix of regression coefficients of \mathbf{X}_m on \mathbf{x}_n given by (Anderson, 2003, Definition 2.5.1)

$$\Gamma_{m|n} = \Sigma_{mn}\Sigma_{nn}^{-1}, \quad m \neq n. \tag{7}$$

2.2. Cumulants

The cumulants of i_d can be calculated in closed form from those of j_d and Eq. (4) by noting that the first cumulant, $\kappa_1(i_d)$, is shift-equivariant, while the others, $\kappa_l(i_d)$ for $l \geq 2$, are shift invariant (Kendall, 1945, §3.13). This leads to

$$\begin{cases} \kappa_1(i_d) &= I(\mathbf{X}_1; \dots; \mathbf{X}_N) + \kappa_1(j_d) \\ \kappa_l(i_d) &= \kappa_l(j_d), \quad l \geq 2. \end{cases} \tag{8}$$

Now, the cumulants of j_d can be easily computed as follows. Using the fact that $|\mathbf{A}| = e^{\text{tr}[\ln(\mathbf{A})]}$ (Higham, 2007), which, for a positive definite matrix, can be expressed as $\ln |\mathbf{A}| = \text{tr}[\ln(\mathbf{A})]$, we have from Eq. (6)

$$\ln E(e^{j_d}) = -\frac{1}{2} \text{tr}[\ln(\mathbf{I}_d - t\Gamma)].$$

For t sufficiently small, we can perform a Taylor expansion of the log function around \mathbf{I}_d (Abramowitz and Stegun, 1972, Eq. 4.1.24), leading to

$$\begin{aligned} \ln E(e^{j_d}) &= \frac{1}{2} \text{tr} \left[\sum_{l=1}^{\infty} \frac{(t\Gamma)^l}{l} \right] \\ &= \sum_{l=1}^{\infty} \frac{t^l}{2l} \text{tr}(\Gamma^l). \end{aligned}$$

Identification with the decomposition of the same function in terms of cumulants (Kendall, 1945, §3.12)

$$\ln E(e^{j_d}) = \sum_{l=1}^{\infty} \kappa_l \frac{t^l}{l!},$$

yields for the cumulants of j_d

$$\kappa_l(j_d) = \frac{(l-1)!}{2} \text{tr}(\Gamma^l). \tag{9}$$

The same result could have been reached by using the fact that j_d is a quadratic function of a multidimensional normal variate \mathbf{x} , as evidenced in Eq. (5), together with the expression of the cumulants of such functions (Magnus, 1986, Lemma 2).

The cumulants of i_d therefore yield

$$\kappa_1(i_d) = I(\mathbf{X}_1; \dots, \mathbf{X}_N),$$

as expected, since the first cumulant is also the mean (Kendall, 1945, §3.14), and, for $l \geq 2$,

$$\kappa_l(i_d) = \frac{(l-1)!}{2} \text{tr}(\Gamma^l).$$

In particular, the variance, which is equal to the second cumulant (Kendall, 1945, §3.14) is given by

$$\text{Var}(i_d) = \kappa_2(i_d) = \sum_{1 \leq m < n \leq N} \text{tr}(\Sigma_{mn}\Sigma_{mm}^{-1}\Sigma_{mn}\Sigma_{nn}^{-1}). \tag{10}$$

In the even more particular case where all subvectors are 1-dimensional, we have

$$\text{Var}(i_d) = \sum_{1 \leq m < n \leq N} \rho_{mn}^2,$$

where ρ_{mn} is the usual correlation coefficient between X_m and X_n .

3. Consequences

We here investigate some consequences of the previous results: the particular case of partitioning into two subvectors, the irrelevance of the variances, and a graphical interpretation.

3.1. Partitioning into two subvectors

In the particular case where $N = 2$, multiinformation boils down to mutual information. The various powers of Γ can easily be calculated, yielding

$$\Gamma^l = \begin{pmatrix} \mathbf{0} & \chi_l \\ \Upsilon_l & \mathbf{0} \end{pmatrix} \quad \text{for } l \text{ odd}$$

and

$$\Gamma^l = \begin{pmatrix} \chi_l & \mathbf{0} \\ \mathbf{0} & \Upsilon_l \end{pmatrix} \quad \text{for } l \text{ even,}$$

where we set

$$\begin{cases} \chi_l &= \underbrace{\Gamma_{1|2}\Gamma_{2|1}\dots}_{l \text{ factors}} \\ \Upsilon_l &= \underbrace{\Gamma_{2|1}\Gamma_{1|2}\dots}_{l \text{ factors}}, \end{cases}$$

with the relationship that $\chi_l = \Gamma_{1|2}\Upsilon_{l-1}$ and $\Upsilon_l = \Gamma_{2|1}\chi_{l-1}$. For l odd, the trace of Γ^l is equal to 0; for l even, it is equal to $\text{tr}(\chi_l) + \text{tr}(\Upsilon_l)$, which can alternatively be expressed as

$$2 \text{tr}(\chi_l) = 2 \text{tr} \left[(\Gamma_{1|2} \Gamma_{2|1})^{\frac{l}{2}} \right]$$

or

$$2 \text{tr}(\Upsilon_l) = 2 \text{tr} \left[(\Gamma_{2|1} \Gamma_{1|2})^{\frac{l}{2}} \right].$$

In particular, the variance of i_d is equal to

$$\text{Var}(i_d) = \kappa_2(i_d) = \text{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}). \quad (11)$$

This quantity, which is a particular case of Eq. (10), was introduced by Jupp and Mardia (1980) as an extension of the classical correlation coefficient in the case of multidimensional variates, with application to directed data (Mardia and Jupp, 2000, §11.2). It is the sum of the squared canonical correlation coefficients between \mathbf{X}_1 and \mathbf{X}_2 (Anderson, 2003, Chap. 12; Jupp and Mardia, 1980).

If we furthermore assume that X_1 is a 1-dimensional vector, the cumulant-generating function yields

$$\ln E(e^{ti_d}) = -\frac{t}{2} \ln(1 - \bar{R}^2) - \ln(1 - t^2 \bar{R}^2),$$

where \bar{R} is the multiple correlation coefficient (Anderson, 2003, § 2.5.2)

$$\bar{R}^2 = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t}{\sigma_1^2}.$$

The cumulants are equal to 0 for l odd and to $\kappa_l(i_d) = (l-1)! (\bar{R}^2)^{\frac{l}{2}}$ for l even. In particular, the variance reads $\text{Var}(i_d) = \bar{R}^2$.

Finally, if both X_1 and X_2 are assumed to be 1-dimensional vectors with correlation coefficient ρ , the cumulant-generating function reads

$$\ln E(e^{ti_d}) = -\frac{t}{2} \ln(1 - \rho^2) - \ln(1 - t^2 \rho^2),$$

with cumulants equal to 0 for l odd and $\kappa_l(i_d) = (l-1)! |\rho|^l$ for l even. In particular, the variance yields $\text{Var}(i_d) = \rho^2$.

3.2. Irrelevance of variances

Mutual information and multiinformation are both quantities that do not depend on the variance coefficients Σ_{kk} 's. This result can be generalized to the cumulant-generating function of i_d , and hence, its distribution, moments and cumulants. Indeed, let

$$\Sigma = \Delta R \Delta$$

be the decomposition of the covariance matrix Σ such that $\Delta = (\Delta_{kl})$ is a diagonal matrix with $\Delta_{kk} = \sqrt{\Sigma_{kk}}$ and $R = (R_{kl})$ is the correlation matrix with $R_{kl} = \Sigma_{kl} / \sqrt{\Sigma_{kk}\Sigma_{ll}}$. Block multiplication of $\Delta R \Delta$ shows that we have $\Sigma_{mn} = \Delta_{mm} R_{mn} \Delta_{nn}$ for any block Σ_{mn} of Σ . Using the fact that $(AB)^{-1} = B^{-1}A^{-1}$ for any two invertible square matrices A and B , we have

$$\Gamma_{m|n} = \Delta_{mm} R_{mn} R_{nn}^{-1} \Delta_{nn}^{-1} = \Delta_{mm} \tilde{\Gamma}_{m|n} \Delta_{nn}^{-1},$$

where $\tilde{\Gamma}_{m|n}$ is the matrix of regression coefficients obtained by application of Eq. (7) to the correlation matrix R instead of the covariance matrix Σ . This result shows that $\Gamma = (\Gamma_{m|n})$ can be factorized into

$$\Gamma = \Delta \tilde{\Gamma} \Delta^{-1},$$

where we set $\tilde{\Gamma} = (\tilde{\Gamma}_{m|n})$. This takes us to

$$\begin{aligned} |I_d - t\Gamma| &= |\Delta(I_d - t\tilde{\Gamma})\Delta^{-1}| \\ &= |I_d - t\tilde{\Gamma}|. \end{aligned}$$

As a conclusion, we have that the cumulant-generating function of a multivariate distribution with covariance matrix Σ is the same as the cumulant-generating function of a multivariate distribution with covariance matrix R , where R is the correlation matrix associated with Σ . This is a translation of the fact that i_d does not depend on the variance coefficients.

3.3. Graphical interpretation

While simplification of the expression for the cumulants through the explicit calculation of Γ^l is challenging in the general case, one can resort to a graphical interpretation of this matrix. Note first that the block (n, m) of Γ^l is given by

$$(\Gamma^l)_{mn} = \sum_{q_1, \dots, q_{l-1}} \underbrace{\Gamma_{mq_1} \dots \Gamma_{q_{l-1}n}}_{l \text{ terms}}$$

and $\text{tr}(\Gamma^l)$ by

$$\text{tr}(\Gamma^l) = \sum_{n=1}^N \text{tr}[(\Gamma^l)_{nn}] = \sum_{n=1}^N \text{tr} \left(\sum_{q_1, \dots, q_{l-1}} \Gamma_{nq_{l-1}} \dots \Gamma_{q_1 n} \right).$$

Consider then the directed and fully connected graph with N nodes $\{1, \dots, N\}$, an arrow from any m to any $n \neq m$ (no self-connections), and corresponding (potentially matrix) weight $\Gamma_{n|m}$. In this graph, a directed loop is a directed path that begins and starts at the same node. It is a k -loop if the directed path is composed of exactly k arrows. For any node n and integer k , let $\mathcal{L}_k^{\rightarrow}(n)$ be the set of all directed k -loops starting and ending at node n and $\mathcal{L}_k^{\rightarrow}$ the set of all k -loops. For any directed path $p = (q_1 \rightarrow \dots \rightarrow q_k)$, define $\tau(p)$ as the trace of the product of the weights along p

$$\tau(p) = \text{tr}(\Gamma_{q_k|q_{k-1}} \dots \Gamma_{q_2|q_1}).$$

With these notations, $\text{tr}(\Gamma^l)$ can be interpreted as the sum of the values taken by τ along all directed l -loops starting and ending at every node of the graph

$$\text{tr}(\Gamma^l) = \sum_{n=1}^N \sum_{p \in \mathcal{L}_l^{\rightarrow}(n)} \tau(p).$$

It can also be seen as the sum of the values taken by τ along all directed l -loops

$$\text{tr}(\Gamma^l) = \sum_{p \in \mathcal{L}_l^{\rightarrow}} \tau(p).$$

See Fig. 1 for an illustration of this interpretation. Note that the fact that 1-loops do not exist is interpreted as $\text{tr}(\Gamma) = 0$.

This graphical interpretation is in particular compatible with a partitioning into two subvectors. In that case, the corresponding directed graph only has two nodes and two arrows, and $\mathcal{L}_l^{\rightarrow} = \emptyset$ for l odd, which is in agreement with the fact that all cumulants of odd order are equal to zero.

4. Discussion

In the present manuscript, we introduced multiinformation density, a random variable that generalizes information density and whose expectation defines multiinformation. We focused on the case of a multivariate normal distribution and derived a closed form for its cumulant-generating function as well as its cumulants. We showed that the cumulant-generating function does not depend on the values taken by the variance coefficients of the covariance matrix, and that the computations required have a simple graphical interpretation. We also considered the special case of a partitioning into two subvectors, showing the relationship between our results and existing quantities.

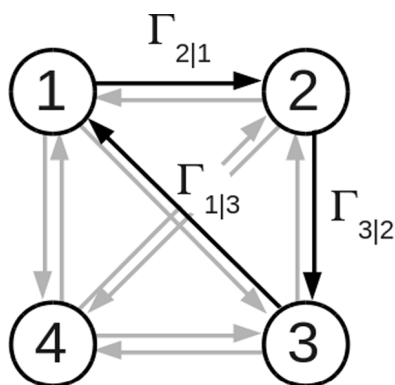


Fig. 1. Graphical interpretation of $\text{tr}(\Gamma^l)$. We consider the case $N = 4$ and $l = 3$. The directed 3-loop $lp = (1 \rightarrow 2 \rightarrow 3 \rightarrow 1)$ is represented with dark arrows. The value of τ on this loop is equal to $\tau(l) = \text{tr}(\Gamma_{1|3}\Gamma_{3|2}\Gamma_{2|1})$. $\text{tr}(\Gamma^3)$ is obtained by summing $\tau(p)$ over all 3-loops.

Interestingly, the results show that the cumulant-generating function of multiinformation density is a function of a specific quantity, namely the block matrix Γ composed of all matrices of regression coefficients corresponding to the partitioning of \mathbf{X} into $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ as defined in Eq. (7). This entails that the probability distribution of multiinformation density as well as all its moments are fully defined by Γ . In particular, it can be shown that mutual information and multiinformation themselves, as expectations, are functions of Γ only, namely (see online supplement, §3)

$$I(\mathbf{X}_1; \dots; \mathbf{X}_N) = -\frac{1}{2} \ln |\mathbf{I}_d + \Gamma|.$$

Going back to our question of knowing whether the multivariate normal distribution family could be parameterized in a natural way with emphasis on its dependence properties, it can be argued that Γ is a good candidate to this aim.

Generalization. The expression of the cumulant-generating function of multiinformation in the case of multivariate normal distributions is quite simple. This simplicity is mostly a consequence of the stability of the multivariate normal family to most operations performed here: (i) the product of the marginals is also a multivariate normal distribution (this is strongly related to the fact that independence and uncorrelatedness are equivalent for multivariate normal distributions); (ii) the ratio of the joint distribution to its marginals takes a simple form that is again closely related to the multivariate normal family; and (iii) the exponentiation of t times the log of this ratio still has the form of a multivariate normal distribution. It would be of interest to determine such a simplicity would still hold in more general settings such as more general distribution families and more general functions of the ratio $f(\mathbf{x})/\prod f_i(\mathbf{x}_i)$.

Regarding the type of family considered, a first step could be to consider the family of multivariate t distributions (Kotz and Nadarajah, 2004). In this case, $f_1(\mathbf{x}_1)$ and $f_2(\mathbf{x}_2)$ are both multivariate t distributions (Kotz and Nadarajah, 2004, §1.10) but $f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)$ itself is not a multivariate t distribution, and the cumulant-generating function of the log of the resulting ratio does not have a simple closed form. Another, further step would be to consider elliptically contoured distributions (Anderson, 2003, §2.7), but again no simplification seems to occur.

Another potential generalization would be obtained by replacing the log function of multiinformation density with a quantity of the form

$$g \left[\frac{f(\mathbf{X})}{\prod f_i(\mathbf{X}_i)} \right],$$

in the spirit of f -divergences (Csiszár, 1963, 1967; Ali and Silvey, 1966; Vajda, 1972). However, by replacing $g(t) = \ln(t)$ with another, more general form, we lose the key property that the log is the inverse of the exponentiation taken to compute the cumulant-generating function, which also simplifies calculations.

Future work. Investigating the level of dependence between variables is still a thorny issue. Such an issue is usually tackled by investigating the properties of mutual information or multiinformation. We believe that considering multiinformation density instead of multiinformation (i.e., the random variable instead of its expectation) could contribute to a better characterization of dependence. For instance, the variance of i_d , given in Eq. (10), could be considered, in addition to or instead of multiinformation, to quantify the presence or absence of dependence. An important point to advance in this direction would be to provide estimators for the quantities obtained here. In the case of two variables, an estimator of this quantity, expressed in Eq. (11), is readily available (Jupp and Mardia, 1980). Its generalization to more than two variables could yield a new tool to investigate mutual independence.

Another point of interest is the investigation of the behavior of multiinformation density in the case of large dimension. For instance, the asymptotic normality of this measure could be of interest. A quick derivation (see online supplement, §4) shows that, in the simple case of a homogeneous correlation matrix and a partitioning into 1-dimensional vectors,

i_d is not asymptotically normal. We hope that tackling this issue in the more general case will help better understand dependences within systems composed of many variables.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.spl.2019.108587>.

References

- Abramowitz, M., Stegun, I.A. (Eds.), 1972. Handbook of Mathematical Functions. In: Applied Math., vol. 55, National Bureau of Standards.
- Ali, S., Silvey, S.D., 1966. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 28, 131–142.
- Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis, third ed. In: Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York.
- Csiszár, I., 1963. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *M. Tud. Akad. Mat. Fiz. Oszt. Közl.* 8, 85–108.
- Csiszár, I., 1967. Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* 2, 229–318.
- Garner, W.R., 1962. Uncertainty and Structure as Psychological Concepts. John Wiley & Sons, New York.
- Higham, N.J., 2007. Functions of matrices. In: Hogben, L. (Ed.), Handbook of Linear Algebra. Chapman & Hall/CRC Press, Boca Raton, Discrete Mathematics and its Applications. chapter 11.
- Joe, H., 1989. Relative entropy measures of multivariate dependence. *J. Amer. Statist. Assoc.* 84, 157–164.
- Jupp, P.E., Mardia, K.V., 1980. A general correlation coefficient for directional data and related regression problems. *Biometrika* 67, 163–173.
- Kendall, M.G., 1945. The Advanced Theory of Statistics. Vol. 1, second ed. Charles Griffin & Co. Ltd., London.
- Kotz, S., Nadarajah, S., 2004. Multivariate T Distributions and their Applications. Cambridge University Press, Cambridge, UK.
- Kullback, S., 1968. Information Theory and Statistics. Dover, Mineola, NY.
- Magnus, J., 1986. The exact moments of a ratio of quadratic forms. *Ann. Économ. Statist.* 4, 95–109.
- Mardia, K.V., Jupp, P.E., 2000. Directional Statistics. In: Wiley Series in Probability and Statistics, Wiley, Chichester.
- Polyanskiy, Y., Wu, Y., 2017. Lecture Notes on Information Theory. <http://www.stat.yale.edu/yw562/ln.html>.
- Studený, M., Vejnarová, J., 1998. The multiinformation function as a tool for measuring stochastic dependence. In: Jordan, M.I. (Ed.), Proceedings of the NATO Advanced Study Institute on Learning in Graphical Models. pp. 261–298.
- Vajda, I., 1972. On the f -divergence and singularity of probability measures. *Period. Math. Hungar.* 2, 223–234.
- Watanabe, S., 1960. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* 4, 66–82.