A Letters Journal Exploring the Frontiers of Physics

December 2009

EPL, **88** (2009) 68003
doi: 10.1209/0295-5075/88/68003

www.epljournal.org

An information-theoretic approach to statistical dependence: Copula information

R. S. Calsaverini[1] and R. Vicente[2(a)]

[1] *Departamento de Física Geral, Instituto de Física, Universidade de São Paulo CP 66318, 05315-970, São Paulo-SP, Brazil*
[2] *Complex Systems, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo 03828-020, São Paulo-SP, Brazil*

received 23 September 2009; accepted 20 November 2009
published online 22 December 2009

PACS 89.70.Cf – Entropy and other measures of information
PACS 89.75.-k – Complex systems
PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – We discuss the connection between information and copula theories by showing that a copula can be employed to decompose the information content of a multivariate distribution into marginal and dependence components, with the latter quantified by the mutual information. We define the information excess as a measure of deviation from a maximum-entropy distribution. The idea of marginal invariant dependence measures is also discussed and used to show that empirical linear correlation underestimates the amplitude of the actual correlation in the case of non-Gaussian marginals. The mutual information is shown to provide an upper bound for the asymptotic empirical log-likelihood of a copula. An analytical expression for the information excess of T-copulas is provided, allowing for simple model identification within this family. We illustrate the framework in a financial data set.

Copyright © EPLA, 2009

Introduction. – Modeling statistical dependence has a pervasive role in science. Information theory provides a unifying framework for ideas from areas as diverse as differential geometry [1], physics [2–4], statistics and telecommunications [5]. From the information-theoretic point of view dependence can be quantified by measuring the distance between a given model defined by a joint probability density $\phi(\mathbf{x})$ and a mean-field model defined by $\phi_0 = \prod_{j=1}^N f_j(x_j)$, where $f_j(x_j)$ are marginal densities $f_j(x_j) = \int \prod_{k\neq j} \mathrm{d}x_k\, \phi(\mathbf{x})$ [6]. The relative entropy given by

$$S\left[\phi \,||\, \phi_0\right] = \int \prod_{j=1}^N \mathrm{d}x_j\, \phi(\mathbf{x}) \log\left(\frac{\phi(\mathbf{x})}{\prod_{j=1}^N f_j(x_j)}\right) \qquad (1)$$

defines a premetric in the space of distributions that can be employed to quantify the degree of dependence in a model, this particular measure is also known as the total correlation or, in the bivariate case, as the mutual information.

The copula theory has been proposed in statistics as an approach for modeling general dependences in multivariate data. A theorem due to Sklar [7] assures that, under very general conditions, for any joint cumulative distribution function (CDF) $F(\mathbf{x}) = \prod_{j=1}^N \int_{-\infty}^{x_j} \mathrm{d}x_j\, \phi(\mathbf{x})$ there is a function $C(\mathbf{u})$ (known as the *copula function*) such that the joint CDF can be written as a function of the marginal CDFs in the form $F(\mathbf{x}) = C[F_1(x_1), \ldots F_N(x_N)]$. The converse is also true: this function couples any set of marginal CDFs to form a multivariate CDF. This provides a convenient picture of the marginals as being responsible for the idiossincratic properties of each variable and the copula function as a description of the dependence between them.

A complete articulation of these two concepts is, however, curiously absent in the literature. In this short contribution we seek to survey the basic ideas connecting these two threads emphasizing the information-theoretic interpretation.

We have organized this letter as follows. In the next section we briefly discuss the idea of measures of dependence that are marginal invariant. We then connect copula theory with mutual information by introducing the concept of copula information and present an analytical prescription to identify a model for bivariate non-Gaussian

(a) E-mail: rvicente@usp.br

dependences within the T-copula family by estimating the mutual information. We briefly comment on general consequences and perspectives in a final section.

Mutual information and copulas. – From this point on we restrict our discussion to bivariate distributions, the multivariate case follows after straightforward adaptations.

Two random variables X and Y are said to be statistically dependent if, and only if, their joint probability density function (PDF) cannot be written as a product of marginal PDFs, that is, if $\phi(x,y) \neq f_x(x) f_y(y)$, where $f_x(x)$ and $f_y(y)$ are marginal densities. A convenient way to quantify statistical dependencies is by evaluating the mutual information defined by

$$I(X,Y) = \int \mathrm{d}x\,\mathrm{d}y\,\phi(x,y)\log\left(\frac{\phi(x,y)}{f_x(x)f_y(y)}\right). \qquad (2)$$

This quantity is a premetric, to say, it is positive and only vanishes in the case of independent variables. By defining the entropy of the distribution of X as $S[f_x] = \int \mathrm{d}x\, f_x(x)\log f_x(x)$ and the average conditional entropy as $S[f_{x|y}] = \int \mathrm{d}y\, f_y(y)\, \int \mathrm{d}x\, f_{x|y}(x)\log f_{x|y}(x)$, where $f_{x|y}(x)$ denotes the conditional probability of X given Y, the identity

$$I(X,Y) = S[f_x] - S[f_{x|y}] \qquad (3)$$

provides an interpretation for the mutual information as the average reduction in the uncertainty in X given knowledge of Y. Alternatively, the mutual information can be regarded as a distance to statistical independence in the space of distributions measured by the relative entropy between the actual joint distribution and the product of marginals $I(X,Y) = S[\phi \parallel f_x f_y]$.

Sklar's theorem asserts that there exists a copula function such that the joint CDF can be written as $F(x,y) = C[F_x(x), F_y(y)]$. We may also regard a copula function as the joint CDF of two uniformly distributed variables u and v, both in the $[0,1]$ interval. Such a pair (u,v) can always be found from any pair of random variables with the substitution $u = F_x(x)$ and $v = F_y(y)$.

To exemplify we can build a joint standard Gaussian with correlation ρ by plugging Gaussian marginal distributions $\Phi(x) = \int_{-\infty}^{x} \frac{\mathrm{d}u}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ into the Gaussian copula defined as

$$C[u,v] = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right), \qquad (4)$$

where $\Phi_\rho(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{\mathrm{d}u\,\mathrm{d}v}{\sqrt{4\pi^2(1-\rho^2)}}\, e^{-\frac{u^2+v^2-2uv\rho}{2(1-\rho^2)}}$.

Clearly X and Y are dependent if, and only if, $C[u,v] \neq uv$. Introducing the copula density as $c[u,v] = \frac{\partial^2}{\partial u \partial v} C[u,v]$, we can decompose the joint probability density as

$$\phi(x,y) = c[F_x(x), F_y(y)] f_x(x) f_y(y) \qquad (5)$$

and observe that statistical dependence would simply imply that $c[u,v] \neq 1$.

Marginal invariant measures. – Two close concepts in statistics are dependence and concordance. While dependence relates only to the functional relationship between two variables, concordance measures whether positive or negative comovement of variables is present. Measures of dependence and concordance are plenty. However, a good dependence (respectively, concordance) measure should [7,8]:

1. be invariant under reparametrizations: $(x,y) \to (q(x), w(y))$, if $q(x)$ and $w(y)$ are monotonous functions (changing sign if one of the reparametrizations is a monotonically decreasing function, in the case of concordance measures),
2. have a unique minimum (a unique zero, in the case of concordance), that can be set to zero with no loss of generality, at $\phi(x,y) = f_x(x) f_y(y)$.

Some authors would also require that a measure of dependence (concordance) should be restricted to the $[0,1]$ ($[-1,1]$) interval. We do not require it here since any real number can be trivially mapped into any interval. Good measures of concordance on the other hand must have a unique zero if X and Y are statistically independent, be invariant under monotonically increasing reparametrizations and change sign if one of the functions of the reparametrization is monotonically decreasing.

With the concept of copula density at hand, these desiderata can be concisely restated as: a measure of dependence must be a functional of the copula density alone (*i.e.* must be independent of marginal densities), with a unique minimum at $c[u,v] = 1$.

The linear correlation for standardized variables $\rho(X,Y) = \int \mathrm{d}x\,\mathrm{d}y\, xy\, \phi(x,y)$ is widely used as a measure of concordance and its absolute value as a measure of dependence. The correlation may be rewritten in terms of copula densities as

$$\rho(X,Y) = \int_{[0,1]^2} \mathrm{d}u\,\mathrm{d}v\, c[u,v]\, F_x^{-1}(u) F_y^{-1}(v). \qquad (6)$$

If X and Y are independent, $c[u,v] = 1$ and consequently $\rho(X,Y) = 0$. However, it is clear that a copula may be chosen such that the linear correlation vanishes even though $c[u,v] \neq 1$. Moreover, $\rho(X,Y)$ is obviously dependent on marginal distributions.

A better alternative for measuring concordance would be the rank correlation, also known as Spearman's ρ defined as

$$\rho_{\mathrm{rank}}(X,Y) = 12 \int_{[0,1]^2} \mathrm{d}u\,\mathrm{d}v\, c[u,v]\, uv - 3. \qquad (7)$$

This measure strictly fulfills concordance measures desiderata. For a Gaussian bivariate distribution, the rank correlation is related to the correlation parameter as:

$$\rho_{\mathrm{rank}}[\Phi_\rho] = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho}{2}\right), \qquad (8)$$

where ρ is the correlation parameter of the Gaussian copula, which is identical to the usual linear correlation *only if the marginals are also Gaussian*. Another measure of dependence that is marginal independent is Kendall's tau defined as

$$\tau(X,Y) = 4\int_{[0,1]^2} \mathrm{d}C\,[u,v]\,C[u,v] - 1. \tag{9}$$

In the case of meta-elliptical distributions [9], that includes Gaussian and T-copulas, Kendall's tau is also related to the the correlation parameter as

$$\tau = \frac{2}{\pi}\sin^{-1}(\rho). \tag{10}$$

In the next section we show that the mutual information also fulfils good dependency measures desiderata, since it is always non-negative, it only vanishes for independent variables and it is a functional of the copula density alone.

Copula information. – Mutual information and copula densities can be connected by plugging eq. (5) into eq. (2), and by performing the simple change of variables $u = F_x(x)$ and $v = F_y(y)$, to conclude that

$$I(X,Y) = \int_{[0,1]^2} \mathrm{d}u\,\mathrm{d}v\,c[u,v]\log(c[u,v]) = -S[c], \tag{11}$$

where $S[c]$ is the differential entropy associated with the $c[u,v]$ distribution, which we will (following [10]) conveniently name the *copula entropy*. Notice that $S[c] \leqslant 0$, as can be shown by considering eq. (5) together with Jensen's inequality, since $-\log(x)$ is a convex function. This simple result shows that mutual information is invariant under arbitrary choices of marginal densities $f_x(x)$ and $f_y(y)$. It is also implied by this connection that using a maximum-entropy principle to choose a copula function given constraints is analogous to assuming the least informative dependence (minimum mutual information) which explains the constraints, which is actually a reasonable principle [11]. This provides yet another interpretation for mutual information: it quantifies the information content of the coupling (copula) functional. From the identity $S[\phi] = S[f_x] + S[f_y] - I(X,Y)$ and eq. (11), we have

$$S[\phi] = S[f_x] + S[f_y] + S[c]. \tag{12}$$

In words: the total information content can be *uniquely* decomposed into the information content in each variable plus the information content on the dependence between them.

Information excess. – When quantifying dependence, it is a common practice to start by measuring linear correlation. In the language we have introduced that is analogous to assuming a Gaussian copula described by a single parameter ρ. However, the notion that this parameter can be measured by the usual linear correlation relies upon the additional assumption that marginals are also Gaussian, as the linear correlation is a measure that also depends on marginals. This particular copula is a very special case as it assumes that the information contained in the dependence between variables is minimal given ρ. This minimal mutual information content in a Gaussian copula is given by [5]

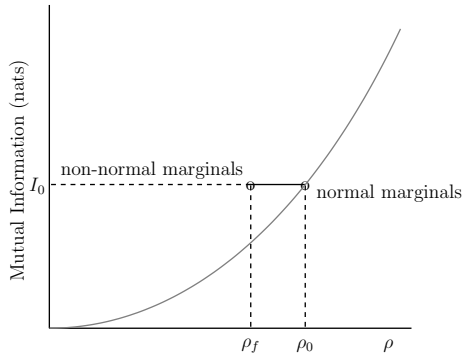

Fig. 1: *Linear correlation is underestimated* in the case of non-Gaussian marginals. If both marginals and copula are Gaussian the joint distribution can be placed over the lower bound for the mutual information. A change in marginals keeping the copula fixed, preserves the mutual information, however correlation estimates are displaced inwards.

$$I_{\mathrm{Gauss}}(\rho) = -\frac{1}{2}\log(1-\rho^2), \tag{13}$$

which can also be written as a function of the *observable* rank correlation using eq. (8). If this assumption of minimal dependence given the parameter ρ fails, an excess of information in the dependence with respect to the Gaussian $I_{\mathrm{excess}} = I(X,Y) - I_{\mathrm{Gauss}}(\rho)$ is observed. An algorithm for efficient estimation of the mutual information $I(X,Y)$ has been proposed in [12] which, together with a good estimate for ρ, provides a diagnostic tool for information excess. The observation of excess means that the dependence cannot be specified by the linear correlation alone even after the identification of non-Gaussian marginals.

If marginals are non-Gaussian neither the mutual information nor the parameter ρ are affected, however, the linear correlation estimate $\rho(X,Y)$ consistently underestimates $|\rho|$. That can be seen by considering the $I(X,Y)$-*vs.*-ρ plane in which the curve described by eq. (13) represents a lower bound for the mutual information as depicted in fig. 1. For a Gaussian copula the parameter ρ is measured by the linear correlation only if marginals are also Gaussian, in this case we can locate a particular joint probability density over the curve of minimal mutual information with a given ρ. Suppose that marginals are changed into non-Gaussian densities. As the copula for the variables is unaltered the mutual information is also unchanged, however, the linear correlation can change. As the curve represents a lower bound for the mutual information given ρ, it is only possible for the linear correlation to change inwards, hence underestimating $|\rho|$. In order

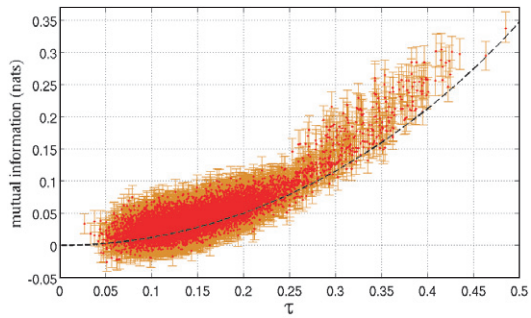

Fig. 2: (Colour on-line) *Mutual Information estimates* following [12] *vs.* Kendall's tau for pairs of series of daily log-returns ($\log[P_{\text{close}}/P_{\text{open}}]$, where P_{close} and P_{open} are, respectively, close and open prices) of 150 stocks composing the S&P500 index over the period from January 2, 1990 to September 16, 2008 (around 4700 samples per series). Bootstrap error bars represent a 90% confidence interval. Note that within this confidence interval a great number of the pairs display a non-zero information excess with respect to the Gaussian copula.

to find ρ correctly we have first to estimate a measure that is marginal invariant, as the rank correlation given by eq. (7), and then employ an inversion relation as eq. (8).

As an applied example, fig. 2 shows estimates for the mutual information obtained by the Kraskov-Stogbauer-Grassberger (KSG) method [12] against Kendall's tau for pairs of series with daily log-returns $r_t = \log[P_{\text{close}}/P_{\text{open}}]$ (where P_{close} and P_{open} are, respectively, close and open prices) of 150 stocks composing the Standard & Poors 500 index (S&P500) over the period from January 2, 1990 to September 16, 2008 (around 4700 points in each series). The error bars have been obtained employing the bootstrap technique [13]. The information excess observed can be traced to time-varying cross-correlations [14] and to dependences between cross-correlations and returns [15] that jointly yield non-Gaussian copulas. Here we have used Kendall's tau as a marginal invariant measure. In the next section we show that this particular marginal invariant plane defined by mutual information and Kendall's tau is sufficient to identify the best T-copula representing non-Gaussian dependences shown by the data.

Copula identification. – Given a data set $\{(x_t, y_t)\}_{t=1}^{T}$ independently sampled from an unknown joint density $\phi(x,y)$, the best approximation $\phi_\theta(x,y)$ within a manifold $\mathcal{F}$, parameterized by θ, can be found by minimizing a sample estimate of the relative entropy [6]:

$$S[\phi \,\|\, \phi_\theta] = \int \mathrm{d}x\,\mathrm{d}y\,\phi(x,y)\log\left[\frac{\phi(x,y)}{\phi_\theta(x,y)}\right]. \tag{14}$$

By considering eq. (5) and performing appropriate variable changes we can write

$$S[\phi \,\|\, \phi_\theta] = S[c \,\|\, c_{\theta_c}] + S[f_x \,\|\, f_x^{\theta_x}] + S[f_y \,\|\, f_y^{\theta_y}], \tag{15}$$

which is just the decomposition (12) in terms of relative entropies. Thus it is reasonably clear that the inference procedure can be implemented by independently minimizing the relative entropy for empirical marginals and copula density. By employing relationship (11), the contribution from the copula in eq. (15) can be further rewritten as

$$S[c \,\|\, c_{\theta_c}] = -L_\infty(\theta_c) - I(X,Y) \geqslant 0, \tag{16}$$

where $L_\infty(\theta_c) = \int_{[0,1]^2} \mathrm{d}u\,\mathrm{d}v\, c[u,v]\log(c_{\theta_c}[u,v])$ is the asymptotic copula log-likelihood. Notice that Jensen's inequality implies that $-L_\infty(\theta_c) \geqslant I(X,Y) \geqslant 0$. Consequently, minimizing $S[c \,\|\, c_{\theta_c}]$ is equivalent to maximizing the likelihood with the mutual information $I(X,Y)$ as a bound.

The estimation of $I(X,Y)$ can be employed to measure the quality of a fit within the chosen family $\mathcal{F}$. In particular, suppose we choose a family such that $L_\infty(\theta_c)$ is known analytically. If we additionally find a family that contains a distribution that saturates the bound, we can use an efficient estimator for the mutual information as [12] to identify the best copula θ_c within $\mathcal{F}$ right away.

In this procedure the identification of the copula is from the start disentangled from the choice of marginals. The T-copula is an interesting choice as the mutual information can be analytically evaluated. The T-copula density is defined in two dimensions as

$$\begin{aligned} c_{\nu,\rho}[u,v] \;=\; & \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\left[\Gamma(\frac{\nu+1}{2})\right]^2\sqrt{1-\rho^2}} \\ & \times \frac{\left[1+\frac{q_\rho(t_\nu^{-1}(u),t_\nu^{-1}(v))}{\nu}\right]^{-\frac{\nu+2}{2}}}{\left[1+\frac{(t_\nu^{-1}(u))^2}{\nu}\right]^{-\frac{\nu+1}{2}}\left[1+\frac{(t_\nu^{-1}(v))^2}{\nu}\right]^{-\frac{\nu+1}{2}}} \end{aligned} \tag{17}$$

with $q_\rho(x,y) = \frac{x^2+y^2-2\rho xy}{1-\rho^2}$ and $t_\nu^{-1}(u)$ denoting the inverse of the distribution function of the univariate Student T density with ν degrees of freedom. It can be shown (see appendix) that the mutual information of a multivariate T-copula can be decomposed as

$$I_{\text{T}}(\rho,\nu) = I_{\text{Gauss}}(\rho) + I_{\text{excess}}(\nu), \tag{18}$$

where, in two dimensions (2D), $I_{\text{Gauss}}(\rho)$ is given by eq. (13). The excess information term only depends on the number of degrees of freedom ν. In 2D it is given by

$$\begin{aligned} I_{\text{excess}}(\nu) \;=\; & 2\log\left(\sqrt{\frac{\nu}{2\pi}}B\left(\frac{\nu}{2},\frac{1}{2}\right)\right) - \frac{2+\nu}{\nu} \\ & +(1+\nu)\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right], \end{aligned} \tag{19}$$

where $B(x,y)$ is the Beta function defined as

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{20}$$

and $\psi(x)$ is the digamma function. Figure 3 shows the T-copula information excess $I_{\text{excess}}(\nu)$ as provided by

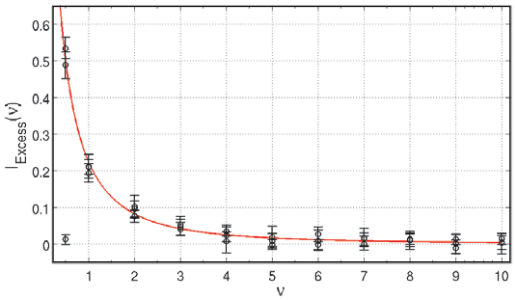

Fig. 3: (Colour on-line) *T-copula information excess.* $I_{\text{excess}}(\nu)$ as provided by eq. (19) (full line). Circles show estimates for 20 runs with T-copulas with known ν for $\rho = 0.1, 0.5, 0.9$ and arbitrary marginals. Error bars represent one standard deviation. Estimates have been computed by employing the KSG method.

eq. (19). The parameter ρ yields the linear correlation in the purely Gaussian case ($\nu \to \infty$) but must be estimated through a marginal independent measure of concordance/dependence in the general case. For T-copulas ρ_{rank} is a function of both ρ and ν that is not known in any simple form. However, in order to identify the appropriate T-copula a simpler alternative is to employ Kendall's tau that is a function of ρ given by eq. (10). We can estimate Kendall's tau and then employ the excess of information in relation to a Gaussian copula to find ν. Figure 3 shows the result of simulations using data sampled from a joint distribution composed by a copula density with known parameters and arbitrary marginals. Going back to fig. 2 the best copula within the manifold of T-copulas can be immediately identified for each point in the mutual information *vs.* Kendall's tau plane.

Conclusions. – The literature on information and copula theories has developed in relative isolation. In this paper we sought to discuss a couple of consequences yield by connections between these two threads.

Copula theory can be employed for factorizing a general joint distribution into marginal fluctuations and a dependence core that is not unique. On the other hand, a combination of copula and information theories provides a unique decomposition in terms of global information content measures. This decomposition yields a simple test of Gaussianity through the estimate of the information excess (a procedure that is simpler than, *e.g.*, [16] or [17]) and also suggests a method for copula identification based on information content matching. This method displays a simple formal equivalence to the usual maximum-likelihood methods (*e.g.*, [18]).

This approach also clarifies the danger of using linear correlation as a measure of dependence for, *e.g.*, portfolio optimization or time series analysis as this measure is bound to underestimate dependence that would be better captured by easily estimated marginal invariant measures [8,12]. Finally, we think that a unified understanding of information and copula theories may be a useful source of new fundamental ideas for the analysis of multivariate data arising from complex physical phenomena.

* * *

This work has been funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under grant 550981/2007. We thank N. Caticha for interesting discussions on maximum-entropy methods.

Appendix: information excess for T-copulas. – In this appendix we present a derivation of the entropy and mutual information of Student T distributions.

The standard Student T distribution in d dimensions is given by

$$p_d(\mathbf{t} \mid \hat{\Sigma}, \nu) = \frac{1}{Z_d(\hat{\Sigma}, \nu)} \left[1 + \frac{\mathbf{t}^{\mathrm{T}} \hat{\Sigma}^{-1} \mathbf{t}}{\nu} \right]^{-\frac{\nu+d}{2}}, \tag{A.1}$$

where ν is a parameter and $\hat{\Sigma}$ is the correlation matrix.

The normalizing prefactor is defined as

$$\frac{1}{Z_d(\hat{\Sigma}, \nu)} = \frac{\Gamma(\frac{d}{2})}{B(\frac{\nu}{2}, \frac{d}{2}) \sqrt{(\pi\nu)^d |\hat{\Sigma}|}} \tag{A.2}$$

with $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ being the Beta function.

For $d = 2$ this simplifies to read

$$p_2(x, y \mid \rho, \nu) = \frac{\Gamma(1 + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2}) \pi\nu \sqrt{1 - \rho^2}} \left[1 + \frac{q_\rho(x, y)}{\nu} \right]^{-(1+\frac{\nu}{2})} \tag{A.3}$$

with $q_\rho(x, y) = \frac{x^2 + y^2 - 2\rho xy}{1 - \rho^2}$.

The differential entropy of a given set of variables $\mathbf{t}$ distributed as $p(\mathbf{t})$ is given by

$$S[p_d] = -\int \mathrm{d}^n\mathbf{t}\, p_d(\mathbf{t}) \log(p_d(\mathbf{t})). \tag{A.4}$$

The mutual information for d dimensions is

$$I(X_1, X_2, \ldots, X_d) = \int p_d(\mathbf{x}) \log \left[\frac{p_d(\mathbf{x})}{\prod_{i=1}^d p_{1;i}(x_i)} \right]. \tag{A.5}$$

For variables with identical marginals (for $p_{1;i}(x) = p_1(x)$ for all $i = 1, 2, \ldots, d$) this can be written in terms of entropies as

$$I(X_1, X_2, \ldots, X_d) = d\, S[p_1] - S[p_d]. \tag{A.6}$$

We employ a "replica trick" [19] to write

$$S[p_d] = -\lim_{n \to 0} \frac{\mathrm{d}}{\mathrm{d}n} \int \mathrm{d}^n\mathbf{t}\, p_d(\mathbf{t})^{n+1}, \tag{A.7}$$

where the limit $n \to 0$ can be regarded as a analytical continuation for a sequence of integers that is known to give sensible results if $p_d(\mathbf{t})$ has a unique extremum. This calculation is not rigorous but is nicely verified by simulations depicted in fig. 3.

We can always simplify this integral by making the transformation $\mathbf{x} = \hat{U}\mathbf{t}$, where $\hat{U}$ is the unitary matrix that diagonalizes $\hat{\Sigma}$. Calling $\mathcal{I}$ the integral in eq. (A.7), we have

$$\mathcal{I} = \frac{1}{Z_d(\hat{\Sigma},\nu)^{n+1}} \int \mathrm{d}^d x \left[1 + \sum_{i=1}^{d} \left(\frac{1}{\lambda_i \nu}\right) x_i^2\right]^{-\frac{1}{2}(n+1)(\nu+d)}, \tag{A.8}$$

where λ_i is the eigenvalue of Σ corresponding to the i-th direction. We can also choose variables $r_i = \sqrt{\left(\frac{1}{\lambda_i \nu}\right)} x_i$ to write

$$\mathcal{I} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \frac{\sqrt{\nu^d |\hat{\Sigma}|}}{Z_d(\hat{\Sigma},\nu)^{n+1}} \int_0^\infty \mathrm{d}r\, r^{d-1} \left[1 + r^2\right]^{-\frac{1}{2}(n+1)(\nu+d)}. \tag{A.9}$$

The above integral is related to the Beta function yielding:

$$\mathcal{I} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \frac{\sqrt{\nu^d |\hat{\Sigma}|}}{2 Z_d(\hat{\Sigma},\nu)^{n+1}} B\left(\frac{1}{2} n(\nu+d) + \frac{\nu}{2}, \frac{d}{2}\right). \tag{A.10}$$

Plugging it into eq. (A.7) and using our definition (A.2) for $Z_d(\hat{\Sigma},\nu)$ gives

$$\begin{aligned} S[p_d] \;=\; & \frac{1}{2} \log\left[(\pi\nu)^d |\hat{\Sigma}|\right] + \log \frac{B\left(\frac{\nu}{2}, \frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\ & + \left(\frac{\nu+d}{2}\right) \left[\psi\left(\frac{\nu+d}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right], \end{aligned} \tag{A.11}$$

where $\psi(x)$ is the digamma function.

The mutual information of the Student d-dimensional distribution can be calculated using eq. (A.6) with the entropy given by eq. (A.11):

$$\begin{aligned} I_d(\hat{\Sigma},\nu) \;=\; & -\frac{1}{2} \log|\,\hat{\Sigma}\,| \\ & + \log\left\{\frac{\left[B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right]^d \Gamma\left(\frac{d}{2}\right)}{\pi^{\frac{d}{2}} B\left(\frac{\nu}{2}, \frac{d}{2}\right)}\right\} - \frac{\nu(d-1)}{2} \psi\left(\frac{\nu}{2}\right) \\ & + \frac{d(\nu+1)}{2} \psi\left(\frac{\nu+1}{2}\right) - \frac{(\nu+d)}{2} \psi\left(\frac{\nu+d}{2}\right). \end{aligned} \tag{A.12}$$

Notice that the only term depending on the correlation matrix $\hat{\Sigma}$ is the mutual information of a Gaussian distribution $I_{\mathrm{Gauss}} = -\frac{1}{2}\log|\hat{\Sigma}|$. The remaining term is the information excess. For $d = 2$ we have

$$I_2(\rho,\nu) = I_{\mathrm{Gauss}} + I_{\mathrm{excess}}, \tag{A.13}$$

with

$$I_{\mathrm{Gauss}} = -\frac{1}{2} \log\left(1 - \rho^2\right) \tag{A.14}$$

and

$$\begin{aligned} I_{\mathrm{excess}} \;=\; & 2\log\left(\sqrt{\frac{\nu}{2\pi}} B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right) - \frac{2+\nu}{\nu} \\ & + (1+\nu)\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right], \end{aligned} \tag{A.15}$$

where we used the fact that $B(x,1) = \frac{1}{x}$ and $\psi(x+1) - \psi(x) = \frac{1}{x}$.

REFERENCES

[1] Amari S.-I. and Nagaoka H., *Methods of Information Geometry* (Oxford University Press) 2007.
[2] Caticha A., arXiv:0808.1260v1 (2008).
[3] Shental O. and Kanter I., *EPL*, **85** (2009) 10006.
[4] Maccone L., *Phys. Rev. Lett.*, **103** (2009) 080401.
[5] Cover T. M. and Thomas J. A., *Elements of Information Theory* (John Wiley & Company) 1991.
[6] Opper M. and Saad D., (Editors), *Advanced Mean Field Theory* (MIT Press) 2001.
[7] Nelsen R. B., *An Introduction to Copulas* (Springer) 2006.
[8] Mari D. D. and Kotz S., *Correlation and Dependence* (Imperial College Press) 2006.
[9] Fang H.-B., Fang K. T. and Kotz S., *J. Multivar. Anal.*, **82** (2002) 1.
[10] Ma J. and Sun Z., arXiv:0808.0845v1 (2008).
[11] Jaynes E. T., *Probability Theory: The Logic of Science* (Cambridge University Press) 2003.
[12] Kraskov A., Stogbauer H. and Grassberger P., *Phys. Rev. E*, **69** (2004) 066138.
[13] Efron B. and Tibshirani R., *An Introduction to the Bootstrap* (Chapman & Hall) 1993.
[14] Conlon T., Ruskin H. J. and Crane M., *Physica A*, **388** (2009) 705.
[15] Longin F. and Solnik B., *J. Finance*, **56** (2001) 649.
[16] Malevergne Y. and Sornette D., *Quant. Finance*, **3** (2003) 231.
[17] Panchenko V., *Physica A*, **355** (2005) 176.
[18] Fernandez V., *Physica A*, **387** (2008) 3615.
[19] Dotsenko V., *Introduction to the Replica Theory of Disordered Statistical Systems* (Cambridge University Press) 2005.