*Communications in Statistics—Theory and Methods*, 36: 1845–1854, 2007

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920601126449

# Dimensionless Measures of Variability and Dependence for Multivariate Continuous Distributions

DANIEL PEÑA[1] AND ANGELIKA VAN DER LINDE[2]

[1]Department of Statistics and Econometrics, Universidad Carlos III, Madrid, Spain
[2]Department of Mathematics and Computer Science, Institute of Statistics, University of Bremen, Germany

*In this article, we suggest dimensionless descriptive measures of multivariate variability and dependence which can be used in comparisons of random vectors of different dimensions. Our work generalizes the measures of scatter and linear dependence proposed by Peña and Rodríguez (2003). The measure of variability we introduce is the rth root of the (transformed) entropy and the measure of dependence is based on the mutual information between the components of an r-dimensional random vector, capturing general stochastic dependence instead of merely linear dependence. We further investigate decompositions of the measure of variability into a measure of scale and a measure of stochastic dependence. The decomposition resulting from independent components provides a representation of variability by the scale of independent components and thus generalizes the explanation of covariance by principal components in classical multivariate analysis. We illustrate our ideas for examples of non Gaussian random vectors.*

**Keywords** Entropy; Independent components; Mutual information.

**Mathematics Subject Classification** Primary 62H20; Secondary 62B10.

## 1. Introduction

The measures of variability and linear dependence that are most often used for multivariate distributions are scalar functions of the covariance and the correlation matrices, as the trace or the determinant. A limitation of these scalar measures is that they cannot be used to compare random variables with different dimensions. For instance, if we have a set of $r$ measurements from a process A and a set of $q$ measurements from another process B we cannot rely on these measures to tell which of the two processes are more variable or more highly dependent

Received November 17, 2005; Accepted October 4, 2006
Address correspondence to Angelika van der Linde, Department of Mathematics and Computer Science, Institute of Statistics, University of Bremen, P.O. Box 330 440, D-28334 Bremen, Germany; E-mail: avdl@math.uni-bremen.de

among their variables. Recently, Peña and Rodríguez (2003) tried to overcome this limitation introducing the Effective Variance and the Effective Linear Dependence as two descriptive measures for multivariate variables both based on the covariance matrix, and they illustrated their use in comparisons of random vectors of different dimensions. These measures are of course not applicable to multivariate distributions for which the covariance matrix or correlation coefficients do not exist and they are restricted to describe linear relationships.

More generally, the entropy and the mutual information have been suggested to capture variability and stochastic dependence of a random vector $X$. Entropies as measures of variability have been investigated for some multivariate distributions like the Gaussian, the Student-$t$, the Poisson, the logistic, or the Weibull distribution (Darbellay and Vajda, 2000; Guerrero-Cusumano, 1996a,b, 1998). The mutual information $I(X) = H(\widetilde{X}) - H(X)$ measures the difference between the entropy $H(X)$ of the joint density of the component variables $X_i$ of $X$ and the entropy $H(\widetilde{X})$ of the product of marginal densities of $X_i$'s corresponding to a random vector $\widetilde{X}$, and it has been proposed as a measure of dependence (Joe, 1989a; Kapur and Dhande, 1986). In this article we extend the ideas inherent in the Effective Variance and the Effective Linear Dependence: we identify these measures as special Gaussian cases of information theoretic measures, Effective Variability and Effective Dependence, which we propose as more general descriptive measures for continuous multivariate distributions. Unlike previous proposals they allow for the comparison of distributions in different dimensions. We introduce these measures as theoretical measures which in practice need to be estimated.

In recent years, a generalization of Principal Component Analysis (PCA), Independent Component Analysis (ICA), has been developed (Hyvärinen et al., 2001; Roberts and Everson, 2001). In its simplest version for a 'whitened' random vector $X$ a transformation $S = AX$ is sought such that $I(S)$ is minimized. Thus the components given in $S$ are required to be as independent as possible, not only uncorrelated. We point out that (scaled) independent components describe the 'volume' of the distribution of $X$, and in this sense 'explain' the (Effective) variability of $X$ like an ellipsoid of concentration 'explains' the volume of a multivariate Gaussian distribution.

Following this outline we first introduce the Effective Variability in Sec. 2 and we describe some of its properties. In Sec. 3 we propose a general measure of stochastic dependence, the Effective Dependence, which is a transformation of the mutual information $I(X)$. In Sec. 4 we study the relation between variability and dependence. We illustrate our approach with examples of non Gaussian distributions in different dimensions in Sec. 5, and Sec. 6 concludes with a brief discussion.

## 2. Variability of a Random Vector

### 2.1. *Basic Definitions*

Let $X = [X_1, \ldots, X_r]^T$ be an $r$-dimensional continuous random vector with $E(X) = \mu_X$, $cov(X) = \Sigma_X$ and density $p$. Let the spectral decomposition of $\Sigma_X$ be given by $\Sigma_X = P\Lambda_X P^T$, where $\Lambda_X$ is the diagonal matrix of eigenvalues, $\lambda_i$, of $\Sigma_X$, and $P$ is the orthogonal matrix of standardized eigenvectors. Let $|\Sigma_X| = \prod_{i=1}^r \lambda_i$ be the generalized variance of $X$. With $var(X_i) = \sigma_i^2$ define $D_{\sigma^2}$ to be the diagonal matrix with entries $\sigma_i^2$. The correlation matrix $R_X$ of $X$ is obtained as $R_X = D_\sigma^{-1}\Sigma_X D_\sigma^{-1}$, where $D_{\sigma^2} = D_\sigma D_\sigma$. Consider further the $r$ component variables with marginal densities $p_i$, and set $\tilde{p}(x) = \prod_{i=1}^r p_i(x_i)$. The random vector with density $\tilde{p}$ is denoted by $\widetilde{X}$, and $E(\widetilde{X}) = \mu_X$, $cov(\widetilde{X}) = D_{\sigma^2}$. The entropy of $X$ is $H(X) = -E_X \ln[p(x)]$ and with $H$ denoting entropy in general, $H(\widetilde{X}) = \sum_{i=1}^r H(X_i)$.

An analogous notation with different subscripts will be used for other random variables.

## 2.2. *Effective Variability*

**Definition 2.1.** It is easy to see that if we compare the generalized variance of $X$ to the generalized variance of $Y = [X_1, \dots, X_r, X_{r+1}]^T$, the latter one can be made arbitrarily larger or smaller than the first one choosing the units of $X_{r+1}$. But if we measure all the variables in the same units the generalized variance cannot increase when introducing new variables. Avoiding this problem in comparison across dimensions, Peña and Rodríguez (2003) suggested to describe the *average* scatter in any direction by the Effective Variance, the $r$th root of the determinant of the covariance matrix, that is

$$V_e(X) := |\Sigma_X|^{1/r}.$$

We propose to generalize this notion by defining the Effective Variability as the $r$th root of a calibrated transformation of the entropy of $X$,

$$v_e(X) := \left(\exp[2H(X) - 2H(N_r)]\right)^{1/r}, \tag{1}$$

where $N_r$ denotes a standard multivariate $r$-dimensional Gaussian vector, $N_r \sim N(0, I_r)$. Hence

$$v_e(X) = \frac{\exp[2H(X)/r]}{2\pi e}. \tag{2}$$

The calibration is motivated by the relation of entropy to covariance for Gaussian distributions: for $X \sim N(\mu_X, \Sigma_X)$, $H(X) = 0.5 \ln[(2\pi e)^r |\Sigma_X|]$ and hence, $v_e(X) = |\Sigma_X|^{1/r} = V_e(X)$.

It is interesting to note that if $X$ is univariate, $r = 1$, $v_e(X)$ reduces to the 'entropy power' introduced by Shannon (1948), which is equal to the variance of a Gaussian distribution with the same entropy as $X$. Similarly, in the multivariate case $v_e(X)$ is equal to the $r$th root of the determinant of the covariance matrix of a Gaussian vector with the same entropy as $X$. For independent components $X_i$ of $X$, $v_e(X)$ thus can be interpreted as an average entropy power of the $X_i$.

For a random vector $X$, the entropy $H(X)$ characterizes the volume of a "typical set". A typical set $A_\varepsilon^{(n)}$ for $X$ is defined as a set of sequences of length $n$ of independent values $x^{(j)}$ of $X$ with empirical entropies close (within $\varepsilon$) to the true entropies (Cover and Thomas, 1991, Ch. 8.6, 9.2). For $n$ sufficiently large it can be shown that $P(A_\varepsilon^{(n)}) \geq 1 - \varepsilon$ and

$$(1 - \varepsilon)\exp[nH(X) - \varepsilon] \leq \mathrm{vol}\big(A_\varepsilon^{(n)}\big) \leq \exp[nH(X) + \varepsilon].$$

In terms of an $r$-dimensional rectangle to describe the range of variability for any single point (element of the sequence), the volume of that rectangle (not that of the typical set) is approximately $\exp[H(X)]$ with an average side length given by

$$(\exp[H(X)])^{1/r} = (2\pi e v_e(X))^{1/2}. \tag{3}$$

This relation can be used to visualize the Effective Variability and the Effective Variance graphically as squares or cubes centered at $E(X)$ with side lengths $(\exp[H(X)])^{1/r}$ and $(2\pi e V_e(X))^{1/2}$, respectively.

**Properties.** The Effective Variability has the following properties:

(a) If $Y = BX$, where $B$ is an orthogonal matrix, then $v_e(Y) = v_e(X)$.

(b) If $Y = BX + m$, where $B$ is a non singular diagonal matrix with entries $b_{ii}$ and $m$ is a vector, then $v_e(Y) = v_e(X)(\prod_{i=1}^{r} b_{ii}^2)^{1/r}$.
(Both properties immediately follow from the transformation theorem for entropies. For a one-to-one affine transformation $Y = BX + m$ one has $H(Y) = H(X) + \ln(|B|_+)$, where $|B|_+$ denotes the absolute value of the determinant of $B$.)

(c) $v_e(X) \geq 0$, and $v_e(X) \to 0$ if $H(X) \to -\infty$.

(d) Let $Z^T = [X^T, Y^T]$ of dimension $r + q$, where $X$ and $Y$ are random vectors of dimensions $r$ and $q$, respectively. We have the relation

$$v_e(Z) \geq v_e(X) \Leftrightarrow v_e(Y \mid X) \geq v_e(X), \tag{4}$$

where $v_e(Y \mid X)$ is based on the conditional entropy $H(Y \mid X) = E_X(H(Y \mid x))$ and $H(Y \mid x)$ is defined using the conditional density $p(y \mid x)$ of $Y$ given $x$. Hence, if the conditional Effective Variability of $Y$ exceeds that of $X$, the Effective Variability of $Z$ does so as well. This property allows for the comparison of random variables in different dimension. If we measure variability using the Effective Variability we will have that by including new variables the variability will increase if the variability of these new variables given the previous ones is larger than the variability of the original variables.

## 3. Measures of Dependence

The determinant of the correlation matrix has been suggested as a global measure of linear dependence. This measure is the ratio of the generalized variances of $X$ and $\widetilde{X}$, where the components of $\widetilde{X}$ have the same variances as those of $X$ but are uncorrelated. This measure captures only linear dependence among the variables $X_i$ and uncorrelatedness implies independence if the variables are jointly Gaussian. We show that a transformation of mutual information, which is a simple modification of the one introduced by Joe (1989a), yields a measure describing general stochastic dependence and can be used in comparing the strength of dependence of random variables in different dimensions.

### 3.1. *Basic Definitions*

The mutual information for $X$ is given by

$$I(X) := H(\widetilde{X}) - H(X). \tag{5}$$

$I(X)$ measures the discrepancy between the joint density capturing dependence and the product of marginal densities representing independence, and hence it is a natural measure of dependence between the components (Joe, 1989a; Kapur and Dhande, 1986).

In order to have a measure in the range [0, 1] the standardizing transformation

$$R_I(X) = (1 - \exp[-2I(X)])^{1/2} \tag{6}$$

was introduced (Joe, 1989a; Guerrero-Cusumano, 1998). It is motivated by a bivariate Gaussian distribution with coefficient of correlation $\rho$, where $I(X) = -(1/2)\ln(|R_X|)$ yields $R_I(X) = \rho$.

### 3.2. *Effective Dependence*

The transformation (6) is not useful for the comparison of random vectors of different dimensions: for instance, if $X = [X_1, \ldots, X_r]^T$ and $Y = [X_1, \ldots, X_r, X_{r+1}]^T$ we always have $|R_Y| \leq |R_X|$ and thus $R_I(Y) \geq R_I(X)$. To overcome this problem we propose a transformation of $I(X)$ different from $R_I(X)$ in (6).

**Definition 3.1.** Let

$$d_e(X) = 1 - \exp[-2I(X)/r] \tag{7}$$

be the Effective Dependence among the components $X_i$ of $X$.

For multivariate normal variables $d_e(X) = 1 - |R_X|^{1/r}$, which is the measure of Effective Linear Dependence proposed by Peña and Rodríguez (2003). In particular, if $X$ is bivariate Gaussian, $d_e(X) = 1 - \sqrt{1-\rho^2}$.

**Interpretations.** Applying the chain rule for $I(X)$, the Effective Dependence can be interpreted as *average measure of dependence between components $X_i$ and $X_1, \ldots, X_{i-1}$*.

$$d_e(X) = 1 - \exp\left[-2\sum_{i=1}^{r} I(X_i \mid X_1, \ldots, X_{i-1})/r\right], \tag{8}$$

where $I(X_i \mid X_1, \ldots, X_{i-1}) = H(X_i) - H(X_i \mid X_1, \ldots, X_{i-1})$ and $H(X_i \mid X_1, \ldots, X_{i-1}) = -E_{X_1,\ldots,X_i}(\ln[p(x_i \mid x_1, \ldots, x_{i-1})])$. Equation (8) is valid for all orderings of $X_i$ and hence also for the average over all permutations. The interpretation of the Effective Linear Dependence as average proportion of explained variability among the component variables is thus generalized to arbitrary continuous distributions. For Gaussian distributions, similar relations are elaborated by Darbellay (1998). Theil and Chung (1988) explore the idea of averaging over permutations for multiple correlation coefficients.

Considering $Z^T = [X^T, Y^T]$, the Effective Dependence is an average of the internal dependence and the cross-dependence. We have

$$\begin{aligned} d_e(Z) &= 1 - \left(\exp[-2I(Z)]\right)^{1/(r+q)} \\ &= 1 - \left(\exp[-2I(Y)]\exp[-2I(X)]\exp[-2I(X, Y)]\right)^{1/(r+q)} \end{aligned} \tag{9}$$

where $I(X, Y)$ denotes the mutual information between $X$ and $Y$.

**Properties.** The Effective Dependence is a standardization of the Effective Variability.

$$\begin{aligned} d_e(X) &= 1 - \left(\exp[2H(X)/r - 2H(\widetilde{X})/r]\right) \\ &= 1 - v_e(X)/v_e(\widetilde{X}), \end{aligned} \tag{10}$$

and the properties discussed for the Effective Variability may be extended to the Effective Dependence.

## 4. Link Between Variability and Dependence: Scale

We re-consider (5) now arranged as

$$H(X) = H(\widetilde{X}) - I(X), \tag{11}$$

and we may interpret the terms in this equation as a decomposition of variability into scale and dependence in turn. As $H(X)$ is invariant under orthogonal transformations, for $Y = BX$, say,

$$H(X) = H(Y) = H(\widetilde{Y}) - I(Y) \tag{12}$$

as well, and "variability" can be decomposed into "scale" and "dependence" in different proportions. An extreme case is obtained if the components of $Y$ are independent, $H(Y) = \sum_{i=1}^{r} H(Y_i)$ and $I(Y) = 0$. If $X$ is multivariate Gaussian, using the spectral decomposition $\Sigma_X = P\Lambda_X P^T$ underlying Principal Component Analysis (PCA), we obtain

$$\begin{aligned} \frac{1}{2}\ln(|\Sigma_X|) &= \frac{1}{2}\ln(|D_{\sigma^2}|) + \frac{1}{2}\ln(|R_X|) \\ &= \frac{1}{2}\ln(|\Lambda_X|). \end{aligned} \tag{13}$$

For non Gaussian variables, PCA has been generalized to Independent Component Analysis (ICA) (Hyvärinen et al., 2001; Roberts and Everson, 2001), where (in the simplest approach) a linear transformation is sought such that the resulting components, $S$, are minimally dependent. Typically, $X$ is "pre-whitened", that is, $S$ is determined as $S = AW$, where $W = \Lambda_X^{-1/2} P^T (X - E(X)) = \Lambda_X^{-1/2}(T - E(T))$, $T$ comprises the principal components of $X$, and $A$ is orthogonal. Then the entropy of a multivariate variable is equal to the entropy of its principal components but these are not Gaussian and not independent. The components of $S = AW$ are minimally dependent but not necessarily independent and thus do not in general coincide with $\widetilde{S}$, the vector corresponding to the product of marginal densities. With

$$H(X) = H(T) = \ln\left(|\Lambda_X^{1/2}|\right) + H(W) = \ln\left(|\Lambda_X|^{1/2}\right) + H(S)$$

we also have

$$H(X) = H\left(\Lambda_X^{1/2} S\right) = H\left(\Lambda_X^{1/2}\widetilde{S}\right) - I(S), \tag{14}$$

and the variability of $X$, $H(X)$, is essentially explained in terms of scale $(H(\Lambda_X^{1/2}\widetilde{S}))$ analogously to PCA in the case of a Gaussian vector $X$.

Using the transformation yielding $v_e(X)$ we can re-write (11) as

$$v_e(X) = v_e(\widetilde{X})(1 - d_e(X)), \tag{15}$$

where the Effective Dependence acts as a shrinkage factor on scale. Note that (15) is equivalent to (10).

## 5. Examples

### 5.1. *Example* 1*: t-Distribution*

Guerrero-Cusumano (1996b, 1998) analyzed standardized returns in local currencies to illustrate the use of entropy as measure of variability and of mutual information as measure of dependence. Daily or weekly returns, called "securities" in portfolio theory, can reasonably assumed to be $t$-distributed. The risk of a portfolio (a linear combination of securities) depends on the variability of the securities. Weakly depending securities are preferred because they offer a potential for risk minimization. For four international indices, the American S & P 500 index (S&P), the Japanese Nikkei Average, the British Financial Times index (FT), and the German DAX index (DAX), six pairs of indices were compared in terms of entropy (Guerrero-Cusumano, 1998) and four trivariate combinations were compared in terms of mutual information (Guerrero-Cusumano, 1996b), each comparison being based on a different data set. We extend the analysis to illustrate comparisons of combinations of securities across dimensions.

Consider a three-dimensional standardized $t$-distribution of $X = [X_1, X_2, X_3]^T$ where $X_1 = S\&P$, $X_2 = DAX$, and $X_3 = FT$, with estimated parameters $X \sim St(0, D, 7.92)$, $D = ((d_{ij}))$, $d_{ii} = 0.747$, $d_{12} = 0.343$, $d_{13} = 0.243$, and $d_{23} = 0.386$. The correlation coefficient between pairs, $\rho$, the Effective Variability and the Effective Dependence attain values listed in the table below. (Our values differ from the corresponding values for mutual information given in Guerrero-Cusumano, 1996b. The difference reflects the difference between the table for a constant $c(\alpha, r)$ in Guerrero-Cusumano, 1996b, which we used and the table for another constant $c_I(a, r)$ in Guerrero-Cusumano, 1996a which we believe not to be correct.)

| $Measure \backslash variables$ | $i \in \{1, 2, 3\}$ | $(1, 3)$ | $(1, 2)$ | $(2, 3)$ | $(1, 2, 3)$ |
|---|---|---|---|---|---|
| $\rho$ | | 0.325 | 0.459 | 0.517 | |
| $d_e$ | | 0.061 | 0.118 | 0.150 | 0.180 |
| $v_e$ | 0.966 | 0.908 | 0.853 | 0.820 | 0.791 |

The order of pairs of variables according to the strength of dependence is the same by $\rho$ and $d_e$ although this latter measure shows larger differences in the dependence among pairs: an increase of 245.9% from (1,3) to (2,3) whereas this increase in correlation is only 159%. Note that the subset $(X_2, X_3) = (DAX, FT)$ has the measure of Effective Dependence closest to that of $X$. Hence, a portfolio composed of DAX and FT might perform similarly to that including also S&P. As the variables are standardized all of them singly have the same value for $v_e$.

Hence the scale term is constant and variability increases as dependence decreases. Therefore, there is no optimal pair of securities exhibiting minimum variability and dependence simultaneously.

### 5.2. *Example 2: Lognormal Distribution*

Consider a stochastic volatility model

$$Z_t = \sigma_t U_t, \quad U_t \underset{iid}{\sim} N(0, 1),$$

where $\sigma_t$ and $U_t$ are independent processes and the conditional variances of $Z_t$ are the volatilities $\sigma_t^2$. We assume the simplest case where $Y_t = \ln[\sigma_t^2]$ follows an AR(1)-structure with mean $m$ and autoregression coefficient $\phi$,

$$Y_t - m = \phi(Y_{t-1} - m) + \varepsilon_t, \quad \varepsilon_t \underset{iid}{\sim} N(0, \tau^2).$$

Stochastic volatility models are used in financial statistics to model time series of returns as an alternative to Garch models. For example, application to a data set of 1,512 daily S&P100 index returns, which was analyzed in the economic literature several times (see Berg et al., 2004 as a recent reference), resulted in parameter estimates $\hat{m} = -9.97$, $\hat{\phi} = 0.98$, and $\hat{\tau} = 0.167$ (based on a Bayesian analysis of a hierarchical model). The volatilities $\sigma_t^2$ describe the financial risk at the market and are therefore of special interest. Under the (alternative non Bayesian) assumption of joint Normality of the vectors $Y_{[t]} = [Y_t, \dots, Y_{t+r}]^T$ commonly applied in maximum likelihood estimation, the volatilities form a log-normal stochastic process. We use the measure of scale $v_e(\tilde{\sigma}_{[t]}^2)$ to describe the average daily financial risk based on $r$ returns. Thus, we refer to the entropy power of the volatilities instead of the variance focusing on highly probable volatilities rather than rare (but potentially disastrous extreme) variances of returns. Obviously the choice of the measure depends on one's attitude towards risk aversion.

Assuming $Y_{[t]} \sim N(m1_r, \Sigma_Y^{(r)})$ with entries $\sigma_Y^{(r)}(i, j) = \tau^2\phi^{|i-j|}/(1-\phi^2)$, we get for the log-volatilities $Y_t$

$$V_e(Y_{[t]}) = \frac{\tau^2}{(1-\phi^2)^{1/r}} = \frac{\tau^2}{(1-\phi^2)} \frac{(1-\phi^2)}{(1-\phi^2)^{1/r}} = V_e(\widetilde{Y}_{[t]})(1 - D_e(Y_{[t]})). \tag{16}$$

For the volatility process $\sigma_t^2$ with log-normal vectors $\sigma_{[t]}^2 = [\sigma_t^2, \dots, \sigma_{t+r}^2]^T$ we have $H(\sigma_{[t]}^2) = H(Y_{[t]}) + rm$ yielding

$$v_e(\widetilde{\sigma}_{[t]}^2) = V_e(\widetilde{Y}_{[t]}) \exp(2m).$$

As $V_e(\widetilde{Y}_{[t]}) = \tau^2/(1-\phi^2)$ is constant (in $r$), $v_e(\tilde{\sigma}_{[t]}^2)$ is also constant depending on the parameters only. Thus weekly or monthly averages of entropy power of daily volatilities $\sigma_t^2$ attain the same value. Using the parameter values given above we obtain $\hat{v}_e(\tilde{\sigma}_{[t]}^2) = 0.7 \exp(2 * 19.94)$. In contrast, the estimated average variance $\widehat{V}_e(\tilde{\sigma}_{[t]}^2) = 2.042 \exp(2 * 19.94)$ is three times higher.

Furthermore, we point out that unlike the Effective Linear Dependence the Effective Dependence $d_e(\sigma_{[t]}) = D_e(Y_{[t]})$ fully captures the structure of dependence of the log-normal process of volatilities induced by (linear) autoregression. In particular, with $\phi < 1$ as in the numerical example the log-volatilities form a stationary time series, where $(1-\phi^2)^{1/r} \uparrow 1$ for $r \to \infty$. Hence, $D_e(Y_{[t]})$ is increasing with $r$, for $r \to \infty$ approaching $\phi^2$, an average level of dependence among components. Thus while on a bivariate level $cov(Y_i, Y_r) < cov(Y_i, Y_j)$ for $i, j < r$, the overall measure of linear dependence $D_e(Y_{[t]})$ captures the accumulation of dependence inherent in the autoregressive structure.

The example also indicates how dimensionless measures might be used to describe and investigate time series by comparisons of sub-intervals over time.

## 6. Discussion

Entropy and variance are well known to be different concepts to describe the variability of a probability distribution. Although formally the Effective Variability generalizes the Effective Variance (as a special Gaussian case of Effective Variability), the entropy measuring the volume of a typical set captures the concentration of probability mass, whereas the variance reflects the actual spread or range of attainable values. Being aware of the difference we do not claim that the Effective Variability is always superior to the Effective Variance. We rather point to the interaction between variability and dependence as formalized in (10) and (15). The Effective Dependence which certainly is more useful than the Effective Linear Dependence is inevitably linked to Effective Variability. For this reason, Independent Components used to explain variablity described in terms of entropy are focused on the prediction of likely observations which may explain their successful applications. Our examples illustrate the potential as well as the limitation of Effective Variability and clearly demonstrate the advantage of the Effective Dependence over the Effective Linear Dependence.

We illustrated our ideas with multivariate continuous distributions for which $H(X)$ and $I(X)$ are known in closed form. In order to use the quantities we propose as descriptive measures given a data set, estimates of $H(X)$ and $I(X)$ are needed. In our examples we obtained such estimates plugging in parameter estimates. More generally, estimating entropy without any distributional assumption is a topic of its own, and particularly in the context of ICA a vast literature has become available. Here we only point to Vasicek (1976), Joe (1989b), Beirlant et al. (1997), Darbellay (1999), Darbellay and Vajda (1999) and the books on ICA mentioned before.

## Acknowledgments

The article was initiated while the second author visited Madrid in 2003, a collaboration funded by grant SEJ2004-03303 of the Ministry of Education, Spain.

## References

Beirlant, J., Dudewicz, E. J., Györfi, L., van der Meulen, E. C. (1997). Nonparametric entropy estimation: an overview. *Int. J. Mathemat. Statist. Sci.* 6:17–39.

Berg, A., Meyer, R., Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *J. Bus. Econ. Statist.* 22:107–120.

Cover, T. M., Thomas, J. A. (1991). *Information Theory.* New York: Wiley.

Darbellay, G. A. (1998). Predictability: an information-theoretic perspective. In: Procházka, A. et al., eds. *Signal Analysis and Prediction.* Boston: Birkhäuser, pp. 249–262.

Darbellay, G. A. (1999). An estimator of the mutual information based on a criterion for independence. *Computat. Statist. Data Anal.* 32:1–17.

Darbellay, G. A., Vajda, I. (1999). Estimation of the information by an adaptive partitioning of the observation space. *IEEE Trans. Information Theor.* 45:1315–1321.

Darbellay, G. A., Vajda, I. (2000). Entropy expressions for multivariate continuous distributions. *IEEE Trans. Information Theor.* 46:709–712.

Guerrero-Cusumano, J.-L. (1996a). An asymptotic test of independence for multivariate $t$ and cauchy random variables with applications. *Information Sci.* 92:33–45.

Guerrero-Cusumano, J.-L. (1996b). A measure of total variability for the multivariate $t$ distribution with applications to finance. *Information Sci.* 92:47–63.

Guerrero-Cusumano, J.-L. (1998). Measures of dependence for the multivariate $t$ distribution with applications to the stock market. *Commun. Statist. Theor. Meth.* 27:2985–3006.

Hyvärinen, A., Karhunen, J., Oja, E. (2001). *Independent Component Analysis.* New York: Wiley.

Joe, H. (1989a). Relative entropy measures of multivariate dependence. *J. Amer. Statist. Assoc.* 84:157–164.

Joe, H. (1989b). Estimation of entropy and other functionals of a multivariate density. *Ann. Instit. Statist. Math.* 41:683–697.

Kapur, J. N., Dhande, M. (1986). On the entropic measures of stochastic dependence. *Indian J. Pure Appl. Math.* 17:581–595.

Peña, D., Rodríguez, J. (2003). Descriptive measures of multivariate scatter and linear dependence. *J. Multivariate Anal.* 58:361–374.

Roberts, S., Everson, R. Eds. (2001). *Independent Component Analysis: Principles and Practice*. Cambridge: Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Technical J.* 27:379–423, 623–656.

Theil, H., Chung, C.-F. (1988). Information-theoretic measures of fit for univariate and multivariate linear regressions. *Amer. Statistician* 42:249–252.

Vasicek, O. (1976). A test for normality based on sample entropy. *J. Roy. Statist. Soc. B* 38:54–59.