Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 213–216
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

Asymptotic properties of a dimension-robust quadratic dependence measure

Sophie Achard

Brain Mapping Unit, University of Cambridge, Downing Site, Cambridge CB2 3EB, UK

Received 23 March 2007; accepted after revision 25 October 2007

Available online 28 January 2008

Presented by Paul Deheuvels

Abstract

The quadratic dependence measure is related to measures used in independence tests, but is derivable, thus suitable for independent component analysis. An adjustable kernel allows to accelerate the convergence of the estimator without affecting the bias. ***To cite this article: S. Achard, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Résumé

Propriétés asymptotiques d'une mesure de dépendance quadratique robuste aux grandes dimensions. La mesure de dépendance quadratique peut-être reliée aux mesures utilisées dans les tests d'indépendance, mais étant de plus dérivable, on peut l'utiliser dans les méthodes d'analyse en composantes indépendantes. Un noyau ajustable permet d'accélérer la convergence de l'estimateur sans pour autant affecter son biais. ***Pour citer cet article : S. Achard, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

1. Introduction

Since the 1950s, there has been a continuous research activity over the definition of measures of dependence: positive functions that are equal to zero if and only if the variables are independent. Hoeffding [8] proposed to define an independence test by comparing the joint cumulative distribution function and the product of the marginal cumulative distribution functions. Then, in the 1970s several authors [10,2,6], have studied independence tests, e.g. by comparing the joint density and the product of the marginal densities. But, in general, these tests are constructed to control the independence of only two variables, and are unsuitable in higher dimensions because of the curse of dimension in estimating the density. In this Note, we study the dependence measure called quadratic dependence [1] whose definition involves an adjustable kernel function. In order to relate this quadratic dependence with other existing dependence measures (e.g. [3,5,9]), we derive two different expressions for it (Section 2). The first one is based on the comparison of the joint characteristic function and the product of the marginal characteristic functions, which allows us to derive asymptotic properties of the estimator. The second one is based on its decomposition as U-statistics, and allows us to prove asymptotic normality and to gain insight on the crucial choice of the kernel bandwidth (Section 3).

E-mail address: sa428@cam.ac.uk.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.10.043

2. A kernel-based characterisation of independence

We introduce a dependence measure which is continuous and derivable, so as to allow convenient minimisation procedures. Let $\mathcal{K}$ be a summable function such that its Fourier transform is different from zero almost everywhere. Then, for any random variables $Y_1, \ldots, Y_K$, the equality of $E[\prod_{k=1}^K \mathcal{K}(y_k - Y_k)]$ and $\prod_{k=1}^K E[\mathcal{K}(y_k - Y_k)]$ for all vectors $(y_1, \ldots, y_K)$ in $\mathbb{R}^K$ is equivalent to the independence of $Y_1, \ldots, Y_K$. Thus, a dependence measure can be obtained by associating this characterisation of dependence with a quadratic measure, as described below.

Definition 1 *(Quadratic dependence).* Let $\mathcal{K}$ be a square-summable kernel function with Fourier transform different from zero almost everywhere. For a set of K random variables $Y_1, \ldots, Y_K$ (with finite variance), we define the quadratic measure of their (mutual) dependence as

$$Q(Y_1, \ldots, Y_K) = \int \left(E\left[\prod_{k=1}^K \mathcal{K}_h\left(y_k - \frac{Y_k}{\sigma_{Y_k}} \right) \right] - \prod_{k=1}^K E\left[\mathcal{K}_h\left(y_k - \frac{Y_k}{\sigma_{Y_k}} \right) \right] \right)^2 \mathrm{d}y_1 \cdots \mathrm{d}y_K, \tag{1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_K)^T$ and σ_{Y_k} is a scale factor, that is, a positive functional of the distribution of Y_k such that $\sigma_{\lambda Y_k} = |\lambda| \sigma_{Y_k}$, for all real constant λ, and $\mathcal{K}_h = \mathcal{K}(x/h)/h$.

Since for any random variables $Y_1, \ldots, Y_K$, $Q(Y_1, \ldots, Y_K)$ is zero if and only if the random variables $Y_1, \ldots, Y_K$ are independent, the function Q is a dependence measure. This follows from the continuity of the characteristic functions and the equivalent expression of Q in terms of the characteristic functions as stated in Lemma 2:

Lemma 2. *Let us define $\psi_{\mathbf{Y}}$ the joint characteristic function of $Y_1, \ldots, Y_K$, ψ_{Y_k} the characteristic function of Y_k, and $\psi_{\mathcal{K}}$ the Fourier transform of $\mathcal{K}$. The quadratic dependence Q can be expressed as a weighted average of the difference between the joint characteristic function and the product of the marginal characteristic functions*:

$$Q(Y_1, \ldots, Y_K) = \frac{1}{(2\pi)^K \prod_{k=1}^K \sigma_{Y_k}} \int \prod_{k=1}^K \left| \sigma_{Y_k} \psi_{\mathcal{K}_h}(\sigma_{Y_k} y_k) \right|^2 \left| \psi_{\mathbf{Y}}(y_1, \ldots, y_K) - \prod_{k=1}^K \psi_{Y_k}(y_k) \right|^2 \mathrm{d}y_1 \cdots \mathrm{d}y_K. \tag{2}$$

This lemma is proved using the Parseval formula, stating that the Fourier transform is unitary. Also, it is easily verified from (1) that the quadratic dependence is invariant by translation and by multiplication by a scalar. The measure (2) has been studied previously [9,5,6], but only in the particular case where $\mathcal{K}$ is a Gaussian kernel and without a scaling factor. When the bandwidth tends to zero and under usual hypotheses for the density and the kernel, Q is equal to the quadratic measure of the difference between the joint density and the product of the marginal densities. In [3], they study the minimum of an estimator of (2) in the context of linear ICA and prove its consistency independently of the choice of a kernel.[1] They point however that the choice of a kernel and especially, variations of its bandwidth, can change dramatically the variance and convergence in moment of the estimators. By defining an independence test based on the quadratic dependence measure, the purpose of the present study is to shed some light on the behaviour of this test in terms of the kernel bandwidth. The quadratic dependence as rewritten in (2) is not easy to estimate because of the multiple integration. The following lemma derives a formula for the quadratic dependence from which a convenient estimator arises. The kernel trick employed for this is specific to this measure, and is a first step to address the problem of the curse of dimension.

Lemma 3. *Let $\mathcal{K}_2$ be the convolution of $\mathcal{K}$ with its mirror, i.e. $\mathcal{K}_2(u) = \int \mathcal{K}(u+v)\mathcal{K}(v)\,\mathrm{d}v$. For a set of K random variables $Y_1, \ldots, Y_K$ (with finite variance), the quadratic measure of their (mutual) dependence is equivalent to*

$$Q(Y_1, \ldots, Y_K) = E\left[\pi_{\mathbf{Y}}(\mathbf{Y})\right] + \prod_{k=1}^K E\left[\pi_{Y_k}(Y_k)\right] - 2E\left[\prod_{k=1}^K \pi_{Y_k}(Y_k) \right],$$

[1] Note also that, in their definition, there is no scaling factor in the weight function, and therefore, no invariance by multiplication by a factor.

$$\pi_{\mathbf{Y}}(\mathbf{y}) = E\left[\prod_{i=1}^{K} \mathcal{K}_{2,h}\left(\frac{y_i - Y_i(n)}{\sigma_{Y_i}}\right)\right], \qquad \pi_{Y_k}(y_k) = E\left[\mathcal{K}_{2,h}\left(\frac{y_k - Y_k(n)}{\sigma_{Y_k}}\right)\right]. \tag{3}$$

The proof of this lemma is obtained by developing the square term and using Fubini's theorem to inverse the integration variables. This lemma shows that Q depends on $\mathcal{K}$ only indirectly through $\mathcal{K}_2$, therefore we can choose $\mathcal{K}_2$ directly without ever considering $\mathcal{K}$. For consistency with its definition, $\mathcal{K}_2$ must be chosen such that its Fourier transform is a positive summable even function. Indeed, its Fourier transform corresponds to $|\psi_{\mathcal{K}}|^2$ where $\psi_{\mathcal{K}}$ is the Fourier transform of a real square-summable function. Moreover, the Fourier transform of $\mathcal{K}_2$ has to be different from zero almost everywhere.

An estimator of Q is defined using expression (3). In the sequel, the observed data will be denoted by $Y_k(n)$, $n = 1, \ldots, N$; $k = 1, \ldots, K$; N being the sample size and the scaling factor $\sigma = (\sigma_{Y_1}, \ldots, \sigma_{Y_K})$ is supposed to be known, that is independent of the sample. Let us remark that (3) involves only the expectation operator E, thus a natural estimator $\widehat{Q}$ of Q can be obtained simply by replacing this operator with the sample average $\widehat{E}$, defined as $\widehat{E}\phi(\mathbf{Y}) = \sum_{n=1}^{N} \phi(\mathbf{Y}(n))/N$, for any function ϕ of K (real) variables. Note that the computational cost of the estimator $\widehat{Q}$ is of order KN^2. As the exact expression of Q is given in terms of the characteristic functions, the estimator $\widehat{Q}$ can alternatively be rewritten in terms of the estimators of the characteristic functions.

3. Asymptotic properties

The asymptotic behaviour of the estimator $\widehat{Q}$ under the hypothesis of dependence of the random variables is given first. Then, using U-statistics, the variance of the estimator $\widehat{Q}$ is computed. Finally, it is shown that the estimator $\widehat{Q}$ converges to a Gaussian distribution.

Lemma 4. *Suppose that the Fourier transform of $\mathcal{K}_2$ is positive, different from zero almost everywhere. Then, under the hypothesis of the dependence of the random variables $Y_1, \ldots, Y_K$, $\lim_{N\to+\infty} \widehat{Q}(Y_1, \ldots, Y_K) > 0$ a.s., for any cumulative distribution function of* $\mathbf{Y}$.

This lemma is proved using that $\sup_{\mathbf{y}\in B} |\hat{\psi}_{\mathbf{Y}}(\mathbf{y}) - \psi_{\mathbf{Y}}(\mathbf{y})| \xrightarrow{\text{a.s.}} 0$ and $\sup_{\mathbf{y}\in B} |\prod_{k=1}^{K} \hat{\psi}_{Y_k}(y_k) - \prod_{k=1}^{K} \psi_{Y_k}(y_k)| \xrightarrow{\text{a.s.}} 0$, $N \to +\infty$ [4]. Unlike the case for the estimation of the density, the estimator $\widehat{Q}$ is unbiased, that is $E[\widehat{Q}] = Q$ [7]. This result is completely independent of the choice of the kernel. Consequently, the bandwidth does not have to assume a specific dependence on the sample size in order to achieve convergence in mean. In particular, the bandwidth does not have to vanish as the sample size tends to infinity. Moreover, the convergence of the estimator of $\widehat{Q}$ does not suffer from the problem of curse of dimensionality. Also we show that the variance of the estimator $\widehat{Q}$ goes to zero for any fixed bandwidth and deduce that $\mathrm{var}(\widehat{Q}) = \mathrm{O}(K^2/N)$. The variance of the estimator depends on the choice of the kernel and its bandwidth, and on the distribution of the observations.

The quadratic dependence measure provides us with an estimator for the evaluation of the dependence between variables. In the following, we construct a hypothesis test of independence based on the quadratic dependence measure. The asymptotic laws under the hypotheses of independence (denoted H_0) and dependence (denoted H_1) are deduced. Finally, it is shown that this hypothesis test of independence is consistent for any choice of the bandwidth.

Law under the hypothesis of independence (denoted H_0):

The estimator $N\widehat{Q}$ follows asymptotically a law of $\gamma\chi^2(\beta)$ where γ and β are $\gamma = V_1/2E_1$ and $\beta = 2E_1^2/V_1$, where $E_1 = \lim_{N\to\infty} NE[\widehat{Q}]$ under H_0, and $V_1 = \lim_{N\to\infty} N\mathrm{var}(\widehat{Q})$ under H_0 [9].

Law under the hypothesis of dependence (denoted H_1):

$\sqrt{N}(\widehat{Q} - Q)$ follows asymptotically a normal law with zero mean and variance in $\mathrm{O}(K^2)$.

Lemma 5. *The independence test defined above is consistent for any choice of the bandwidth: Given α, the level of significance, we define q_α the smallest number satisfying the inequality $P_{\mathrm{H}_0}(\widehat{Q} > q_\alpha) = 1 - F_{\gamma\chi^2(\beta)}(Nq_\alpha) \leqslant \alpha$. Then, the power of the test $1 - P_{\mathrm{H}_1}(\widehat{Q} < q_\alpha)$ tends to 1 as N goes to infinity.*

In addition, the power of the test admits a lower bound:

$$1 - P_{\mathrm{H}_1}(\widehat{Q} < q_\alpha) = P_{\mathrm{H}_1}(\widehat{Q} > q_\alpha) > 1 - \mathrm{var}(\widehat{Q})/(q_\alpha - Q). \tag{4}$$

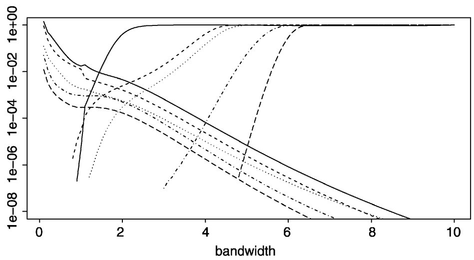

Fig. 1. Variance (decreasing) and type II error (increasing) for different sample sizes. —, $N = 100$, - - -, $N = 200$, $\cdots$, $N = 400$, - · -, $N = 800$, – – –, $N = 1600$.

This lemma is proved using the Chebychev inequality and the asymptotic properties of the variance. Note that the lower bound in (4) is not sharp, as is illustrated in Fig. 1.

It is not necessary to choose the bandwidth so as to make a tradeoff between the bias and the variance, since the asymptotic bias of the estimator tends to zero when the size of the sample N goes to infinity without any constraint on the bandwidth. As a result, the bandwidth can rather be adjusted in a tradeoff between the minimisation of the variance and the minimisation of the type II error of the test. With the expression of the variance given above, it is clear that when the bandwidth h of the kernel increases, the variance of the estimator will decrease. But, as the bandwidth h increases, the type II error of the test is expected to increase. Indeed, the asymptotic power of the test is defined by $1 - P_{\mathrm{H}_1}(\widehat{Q} < q_\alpha) = 1 - \Phi((q_\alpha - Q)\sqrt{N}/\sigma)$ where for a given α, q_α verifies $P_{\mathrm{H}_0}(\widehat{Q} > q_\alpha) = 1 - F_{\gamma\chi^2(\beta)}(Nq_\alpha)$. Fig. 1 illustrates the behaviour of optimal choices of the bandwidth depending on the size of the sample. In this figure, we observe that for a given bandwidth the convergence of the variance of the estimator is slow (of order $1/N$). But, if the bandwidth is adjusted to make a tradeoff between the variance and the power of the test, the convergence rate is tremendously increased. Future work has to be done to quantify this increase and to propose a computational rule to optimise the bandwidth.

4. Conclusion

The quadratic dependence measure is revisited and its asymptotic properties are demonstrated. The convergence rate in terms of the variance is of order $1/N$, and the power of the test defined by this measure converges to one with a rate of $1/N$ at least, N being the sample size. The introduction of a kernel frame in the definition of the quadratic dependence measure enables us to propose an efficient estimator of computational cost of order KN^2, with K the dimension of the problem. This kernel is adjusted with a bandwidth whose choice does not affect the bias. This differs with the case of the estimation of density. Because of this property, the bandwidth can be chosen in terms of the sample size N, so as to increase the convergence rate of the variance and of the power of the test rather than debiasing the estimator.

Acknowledgement

The author would like to thank D.-T. Pham for pointing out some theoretical issues.

References

[1] S. Achard, D.T. Pham, C. Jutten, Quadratic dependence measure for non linear blind sources separation, in: Proc. Int. Workshop ICA2003, April 2003, pp. 263–268.
[2] J.R. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution functions, Ann. Math. Stat. 32 (1961) 485–498.
[3] A. Chen, P.J. Bickel, Consistent independent component analysis and prewhitening, IEEE Trans. Signal Processing 53 (10) (2005) 3625–3632.
[4] S. Csörgő, Limit behaviour of the empirical characteristic function, Ann. Probab. 9 (1) (1981) 130–144.
[5] J. Eriksson, V. Koivunen, Characteristic function based independent component analysis, Signal Processing 83 (2003) 2195–2208.
[6] A. Feuerverger, A consistent test for bivariate dependence, Int. Statist. Rev. 61 (3) (1993) 419–433.
[7] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Stat. 19 (1948) 293–325.
[8] W. Hoeffding, A non-parametric test of independence, Ann. Math. Stat. 19 (1948) 546–557.
[9] A. Kankainen, Consistent testing of total independence based on empirical characteristic functions, PhD thesis, University of Jyväskylä, 1995.
[10] M. Rosenblatt, A quadratic measure of deviation of two-dimensional density estimates and a test of independence, Ann. Statist. 3 (1) (1975) 1–14.