PROPERTIES OF THE EMPIRICAL CHARACTERISTIC FUNCTION AND ITS APPLICATION TO TESTING FOR INDEPENDENCE

Noboru Murata

Waseda University
Department of Electrical, Electronics, and Computer Engineering
3-4-1 Ohkubo, Shinjuku, Tokyo 169-8555, JAPAN

ABSTRACT

In this article, the asymptotic properties of the empirical characteristic function are discussed. The residual of the joint and marginal empirical characteristic functions is studied and the uniform convergence of the residual in the wider sense and the weak convergence of the scaled residual to a Gaussian process are investigated. Taking into account of the result, a statistical test for independence against alternatives is considered.

1. INTRODUCTION

Independent component analysis (ICA,[1]) deals with a problem of decomposing signals which are generated from an unknown linear mixture of mutually independent sources, and is applied to many problems in various engineering fields, such as source separation problem, dimension reduction, and data mining. ICA is often compared with principle component analysis (PCA), which is a method of multivariate data analysis based on the covariance structure of signals with the Euclidean metric. However, in the ICA framework, the notion of independence, which is not affected by the metric of signals, plays a central role. Supported by the recent improvement of computer systems, which allows us to handle massive data and to estimate computationally tedious higher order information within reasonable time, the concept of ICA gives a new viewpoint of using the mutual information instead of the covariance to the multivariate data analysis. It is also pointed out that, in the context of learning, self-organization of information transformations based on infomax criteria in neural networks is closely related with ICA ([2], and so on).

A simple setting of ICA is as follows. Let $\boldsymbol{S}(t) = (S_1(t), \ldots, S_n(t))^T$ ($t = 0, 1, 2, \ldots$) be a source signal, which consists of zero mean n independent random variables. We assume that each component is an i.i.d. sequence or a strong stationary ergodic sequence, and at most only one component is subject to a Gaussian distribution and the others are non-Gaussian. Let $\boldsymbol{X}(t) = (X_1(t), \ldots, X_m(t))^T$ be an observation and we assume that it is produced as a linear transformation of the source signal by an unknown matrix $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, that is

$$\boldsymbol{X}(t) = A\boldsymbol{S}(t) = \boldsymbol{a}_1 S_1(t) + \cdots + \boldsymbol{a}_n S_n(t). \tag{1}$$

Here the problem is to recover the source signal $\boldsymbol{S}(t)$ and estimate the matrix A from the observation without any knowledge about them, that is, find a matrix B such that the elements of a recovered signal $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_n(t))^T$, which is determined by a certain matrix B as

$$\boldsymbol{Y}(t) = B\boldsymbol{X}(t), \tag{2}$$

are mutually independent. In other words, find a decomposition of observation $\boldsymbol{X}(t)$,

$$\boldsymbol{X}(t) = \tilde{\boldsymbol{a}}_1 Y_1(t) + \cdots + \tilde{\boldsymbol{a}}_n Y_n(t) \tag{3}$$

where elements of $\boldsymbol{Y}(t)$ are mutually independent. Ideally the matrix B is an inverse (or generalized inverse) of the matrix A. However, as well known, the order and amplitude of elements of the recovered signal are ambiguous.

In general, ICA algorithms are implemented as minimization or maximization of a certain contrast function which measures the independence among elements of the recovered signal $\boldsymbol{Y}(t)$. Motivated by mainly stability of algorithms and speed of the convergence, many contrast functions are proposed, and typical ones are as follows. Let $P_{Y_1 \ldots Y_n}(y_1, \ldots, y_n)$ and $P_{Y_i}(y)$ be the joint probability density and the marginal probability density of the recovered signal, respectively. The definition of independence is stated as

$$P_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = P_{Y_1}(y_1) \cdots P_{Y_n}(y_n). \tag{4}$$

Therefore, a certain statistical distance between the joint and the marginal distributions can be a contrast function, such as the Kullback-Leibler divergence and the Hellinger distance. Since these distances are defined with probability distributions, we have to approximate probability distributions by kernel methods or polynomial expansion based on moments or cumulants, in order to calculate them by using only observations, i.e. by using empirical distributions. In practice, the Kullback-Leibler divergence is often used, and in this case entropy or differentials of entropy of the recovered signals have to be estimated utilizing such as Gram-Charlier expansion or non-linear function approximations ([3] and so on).

There are also some methods using higher order statistics in which shape of probability distributions are not exactly assumed. A method of blind separation by Jutten and Herault [4] is one of this type. In the simplest case, they assume only the symmetry of the distributions and use the condition that

$$E(Y_i^3 Y_j) = 0,\ i,j = 1,\dots,n,\ i \neq j. \tag{5}$$

Also cumulants, in particular kurtosis, are often used in contrast functions ([5] and so on).

In another approach, the correlation functions are used in contrast functions to utilize the time structure of the signal ([6] and so on). In this case, the assumption on the source signals may be relaxed to be weakly stationary. From the assumption of independence, the correlation function matrix of the source signal is a diagonal matrix

$$E\left(\boldsymbol{S}(t)\boldsymbol{S}(t+\tau)^T\right) = \begin{pmatrix} R_{S_1}(\tau) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & R_{S_n}(\tau) \end{pmatrix}, \tag{6}$$

at any time delay τ, where $R_{S_i}(\tau)$ denotes the auto-correlation function of the signal $S_i(t)$. Therefore, a desired matrix B is a matrix which simultaneously diagonalizes the correlation function matrices of the recovered signal at any time delays. In practical calculation, this strategy is reduced as an approximated simultaneous diagonalization of several delayed correlation matrices under the least mean square criterion.

So far based on various contrast functions, many algorithms are proposed. Since each method has its own advantages and disadvantages, in practic it is difficult to know which algorithm should be used in advance. We sometimes need to evaluate the validity of the result of an algorithm, or compare the results of different algorithms from the viewpoint of independence of reconstructed signals. Therefore, it is important to consider methods of test for independence. In this paper, we propose to use the empirical characteristic function, and investigate basic properties with numerical experiments.

2. PROPERTIES OF EMPIRICAL CHARACTERISTIC FUNCTION

The characteristic function of a probability distribution F is defined as the Fourier-Stieltjes transform of F,

$$c(t) = E\left(\exp(itX)\right),$$

where E denotes expectation with respect to the distribution F. From the well-known property of the Fourier-Stieltjes transform, the characteristic function and the distribution correspond one to one. Therefore, closeness of two distributions can be restated in terms of closeness of their characteristic functions. Moreover the characteristic function is overall bounded, that is

$$\left|E\left(\exp(itX)\right)\right| \leq E\left(\left|\exp(itX)\right|\right) = 1.$$

These uniqueness and boundedness are specific properties of the characteristic function. The moment generating function which is the Laplace transform of the probability distribution is similar to the characteristic function, but it is not bounded, and there are some pathological examples of distributions which, for instance, have all degrees of moments, but do not have the moment generating functions. Particularly the boundedness is an advantage for stability of numerical calculations.

The empirical characteristic function is defined as

$$c_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itX_j), \tag{7}$$

where X_j,$(j = 1,\dots,n)$ is an i.i.d. sequence from the distribution F. Obviously, the empirical characteristic function is directly calculated from the empirical distribution, We should, however, note that all the calculation is done in the complex domain.

The following two results by Feuerverger and Mureika[7] are quite important to know the properties of convergence of the empirical characteristic function to the true characteristic function.

Theorem 1 (Feuerverger and Mureika[7]) *For $T < \infty$,*

$$P\left\{\lim_{n\to\infty} \sup_{|t|\leq T} |c_n(t) - c(t)| = 0\right\} = 1 \tag{8}$$

holds.

Theorem 2 (Feuerverger and Mureika[7]) *Let $Y_n(t)$ be a stochastic process that is a residual of the empirical characteristic function and the true characteristic function:*

$$Y_n(t) = \{c_n(t) - c(t)\}\sqrt{n}, \quad (-T \leq t \leq T). \tag{9}$$

As $n \to \infty$, $Y_n(t)$ converges to a zero mean complex Gaussian process $Y(t)$ satisfying $Y(t) = Y^*(-t)$ and

$$E\{Y(t)Y(s)\} = c(t+s) - c(t)c(s), \tag{10}$$

where * denotes complex conjugate.

As $n \to \infty$, by the strong law of large numbers, the empirical characteristic function converges almost surely to the true characteristic function pointwisely with respect to t. Furthermore, the above theorems respectively claim that the convergence with respect to t is uniform in the wider sense, and the residual, as a stochastic process, converges weakly to a complex Gaussian process. Feuerverger and Mureika have mainly discussed a test for symmetry of the distribution, but also mentioned about test for fitness, test for independence, applications to estimating parameters, and so on. In practice, a studentized version is often used and the properties of the studentized characteristic function

$$\tilde{c}_n(t) = c_n(t/S) \tag{11}$$

where

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2, \quad \bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \tag{12}$$

are also investigated by Murota and Takeuchi[8]. Similarly, the residual converges to a complex Gaussian process.

We extend the above result to the case of the empirical characteristic function of two independent random variables. Let (X_j, Y_j), $(j = 1, \ldots, n)$ be mutually independent i.i.d. sequences, and $c^X(t)$, $c^Y(s)$, and $c^{XY}(t,s)$ be the characteristic functions of X_j and Y_j, and their joint distributions, respectively. Note that in the two dimensional case, the characteristic function is defined as

$$c^{XY}(t,s) = E(\exp(itX + isY)), \tag{13}$$

and from the assumption of independence between X and Y, it is rewritten as

$$c^{XY}(t,s) = c^X(t)c^Y(s). \tag{14}$$

The empirical characteristic functions are defined as

$$c_n^X(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(itX_j), \tag{15}$$

$$c_n^Y(s) = \frac{1}{n}\sum_{j=1}^{n}\exp(isY_j), \tag{16}$$

$$c_n^{XY}(t,s) = \frac{1}{n}\sum_{j=1}^{n}\exp(itX_j + isY_j). \tag{17}$$

By the strong law of large numbers, the above empirical characteristic functions converge almost surely to the true characteristic functions $c^X(t)$, $c^Y(s)$ and $c^{XY}(t,s)$ pointwisely with respect to t and s, respectively. We can show following two theorems.

Theorem 3 *For $T < \infty$,*

$$P\left\{\lim_{n\to\infty}\sup_{|t|,|s|\le T}|c_n^{XY}(t,s) - c_n^X(t)c_n^Y(s)| = 0\right\} = 1 \tag{18}$$

holds.

Theorem 4 *Let $Z_n(t,s)$ be a two-dimensional complex stochastic process:*

$$Z_n(t,s) = \{c_n^{XY}(t,s) - c_n^X(t)c_n^Y(s)\}\sqrt{n}, \quad (-T \le t, s \le T). \tag{19}$$

As $n \to \infty$, $Z_n(t,s)$ converges weakly to a zero mean complex Gaussian process $Z(t,s)$ satisfying $Z(t,s) = Z^*(-t,-s)$ and

$$E\{Z(t,s)Z(t',s')\} = \{c^X(t+t') - c^X(t)c^X(t')\}\{c^Y(s+s') - c^Y(s)c^Y(s')\}. \tag{20}$$

We have to be careful about the fact that $c_n^{XY}(t,s)$ and $c_n^X(t)c_n^Y(s)$ are not independent, which gives the essential difference from Feuerverger and Mureika's theorems. For the proof of Theorem 3, the equality

$$\begin{aligned} c_n^{XY}(t,s) - c_n^X(t)c_n^Y(s) &= \{c_n^{XY}(t,s) - c^{XY}(t,s)\} \\ &- c^X(t)\{c_n^Y(s) - c^Y(s)\} - \{c_n^X(t) - c^X(t)\}c^Y(s) \\ &- \{c_n^X(t) - c^X(t)\}\{c_n^Y(s) - c^Y(s)\}. \end{aligned}$$

and the continuity theorem of the characteristic function are applied. Knowing that

$$Z_n^{XY}(t,s) = \{c_n^{XY}(t,s) - c^{XY}(t,s)\}\sqrt{n}, \tag{21}$$

$$Z_n^X(t) = \{c_n^X(t) - c^X(t)\}\sqrt{n}, \tag{22}$$

$$Z_n^Y(s) = \{c_n^Y(s) - c^Y(s)\}\sqrt{n} \tag{23}$$

converge pointwisely to Gaussian distributions satisfying

$$E\{Z_n^{XY}(t,s)Z_n^{XY}(t',s')\} = c^X(t+t')c^Y(s+s') - c^X(t)c^X(t')c^Y(s)c^Y(s'), \tag{24}$$

$$E\{Z_n^X(t)Z_n^X(t')\} = c^X(t+t') - c^X(t)c^X(t'), \tag{25}$$

$$E\{Z_n^Y(s)Z_n^Y(s')\} = c^Y(s+s') - c^Y(s)c^Y(s'), \tag{26}$$

$$E\{Z_n^{XY}(t,s)Z_n^X(t')\} = \{c^X(t+t') - c^X(t)c^X(t')\}c^Y(s), \tag{27}$$

$$E\{Z_n^{XY}(t,s)Z_n^Y(s')\} = c^X(t)\{c^Y(s+s') - c^Y(s)c^Y(s')\}, \tag{28}$$

$$E\{Z_n^X(t)Z_n^Y(s)\} = 0 \tag{29}$$

the proof of Theorem 4 basically follow Feuerverger and Mureika's.

Note that $Z_n(t,s)$ consists of only the empirical characteristic functions, which can be calculated from observations. That means we can calculate $Z_n(t,s)$ without any knowledge about the underlying distributions of X and Y.

Also the multidimensional versions for 3 or more random variables can be discussed in a straightforward way.

3. APPLICATION TO TESTING FOR INDEPENDENCE

According to the results in the previous section, we construct a test statistics with

$$Z_n(t,s) = \{c_n^{XY}(t,s) - c_n^X(t)c_n^Y(s)\}\sqrt{n}.$$

Roughly speaking, the procedure for testing consists of evaluating how $Z_n(t,s)$ is close to 0 with a certain distance.

L^p-norm, L^∞-norm and Sobolev-norm may be candidates as a distance function, however, evaluation on the whole R^2 is not necessarily plausible, because Theorems 1 and 3 claim that the residual converges uniformly only on a bounded domain. Actually, the function $Z_n(t,s)$ approaches to 1 arbitrarily often as $t,s \to \infty$ and does not converge uniformly on R or R^2. Therefore, at least from the practical point of view, the distance should be evaluated as

$$\int_{R^2} ||Z_n(t,s)|| \hat{w}(t,s) dtds$$

with a weight function $\hat{w}(t,s)$ whose support is bounded. As a special case, the L^2-norm is used and $\hat{w}(t,s)$ can be decomposed as $\hat{w}(t,s) = |\hat{k}(t,s)|^2$ using a function $\hat{k}(t,s)$ whose inverse Fourier transform is $k(x,y)$. The norm is then rewritten as

$$\begin{aligned}\int |Z_n(t,s)\hat{k}(t,s)|^2 dtds &= \int \Big| \frac{1}{n}\sum_i k(x - X_i, y - Y_i) \\ &\quad - \frac{1}{n^2}\sum_{i,j} k(x - X_i, y - Y_j)\Big|^2 dxdy.\end{aligned}$$

In addition, if the function k is a probability density function, the above distance is interpreted as the L^2-norm of the difference between joint and marginal probability density functions estimated by the kernel method. As another special case, Sobolev-norm weighted around the origin is almost equivalent to the method of using higher order statistics (moments and cumulants), because the differential coefficients at the origin of the empirical characteristic function give the information of cumulants.

In any case, these norms are not plausible in practice, because of the computational cost of the multiple integrals with respect to t and s, and weight functions should be carefully chosen with quantitative evaluation of the convergence using, for instance, cross-validation.

Here we try to reduce the computational cost taking account of the properties of the statistics $Z_n(t,s)$. We should note that the variance of $Z_n(t,s)$ vanishes on the coordinate axes of (t,s), that is, $t = 0$ or $s = 0$, and is bounded at the other points. Also from the covariance structure of the neighborhood, we know that $Z_n(t,s)$, as a stochastic process, changes continuously and smoothly (Fig. 5, see for example). As discussed in Murota and Takeuchi[8], we evaluate $Z_n(t,s)$ at one point or several points on a certain region of (t,s) taking account of the continuity. Since $Z_n(t,s)$ almost obeys a Gaussian distribution if n is sufficiently large, we can adopt the L^2-norm of $Z_n(t,s)$ normalized with the variance

$$T(t,s) = (\mathrm{Re}Z_n(t,s)\ \mathrm{Im}Z_n(t,s))\,\Sigma^{-1}\begin{pmatrix}\mathrm{Re}Z_n(t,s)\\ \mathrm{Im}Z_n(t,s)\end{pmatrix} \quad (30)$$

where the elements of the matrix Σ are given by

$$\begin{aligned}\Sigma_{11} &= \mathrm{Var}(\mathrm{Re}Z_n(t,s)) \\ &= \frac{1}{2}\{\mathrm{Re}\mathrm{Var}(Z_n(t,s)) + \mathrm{Cov}(Z_n(t,s), Z_n^*(t,s))\}, \\ \Sigma_{12} &= \Sigma_{21} = \mathrm{Cov}(\mathrm{Re}Z_n(t,s), \mathrm{Im}Z_n(t,s)) \\ &= \frac{1}{2}\mathrm{Im}\mathrm{Var}(Z_n(t,s)), \\ \Sigma_{22} &= \mathrm{Var}(\mathrm{Im}Z_n(t,s)) \\ &= \frac{1}{2}\{-\mathrm{Re}\mathrm{Var}(Z_n(t,s)) + \mathrm{Cov}(Z_n(t,s), Z_n^*(t,s))\}.\end{aligned}$$

Note that $Z_n^*(t,s) = Z_n(-t,-s)$ holds, and then Σ can be asymptotically calculated as

$$\begin{aligned}\mathrm{Var}(Z_n(t,s)) &\sim \{c_n^X(2t) - c_n^X(t)^2\}\{c_n^Y(2s) - c_n^Y(s)^2\} \\ \mathrm{Cov}(Z_n(t,s), Z_n^*(t,s)) &\sim \{1 - |c_n^X(t)|^2\}\{1 - |c_n^Y(s)|^2\}.\end{aligned}$$

This test statistics obeys the χ^2 distribution with 2 degrees of freedom, hence *p*-values can be evaluated as, for example,

$$\begin{aligned}Prob\{T(t,s) \le 4.6052\} &= 0.9 \\ Prob\{T(t,s) \le 5.9915\} &= 0.95 \\ Prob\{T(t,s) \le 9.2103\} &= 0.99\end{aligned}$$

under the null hypothesis that two random variables are mutually independent.

4. NUMERICAL EXPERIMENTS

4.1. Properties of test statistics

First, we check the properties of the test statistics and their p-values discussed in the previous section. In each run, we generate independent random variables $\{X_i; i = 1,\ldots,500\}$, $\{Y_i; i = 1,\ldots,500\}$ subject to the beta distribution $\beta(1,2)$ and calculate the test statistics $T(1,1)$. Figures 1 and 2 shows the histograms of the test statistics and p-values made from 2000 runs. Theoretically the test statistics obeys χ^2

distribution with 2 degrees of freedom and the p-value obeys uniform distribution on $[0, 1]$. Comparison of experimental and theoretical cumulative distributions of the test statistics in Figure 3 supports our analysis.

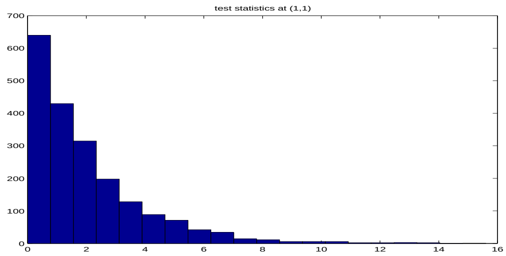

Fig. 1. Histogram of test statistics.

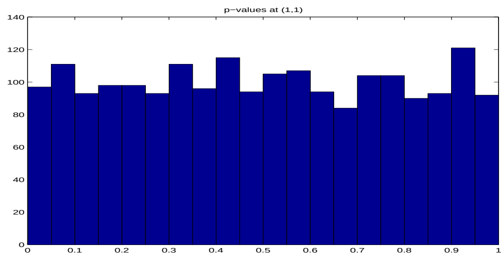

Fig. 2. Histogram of p-values.

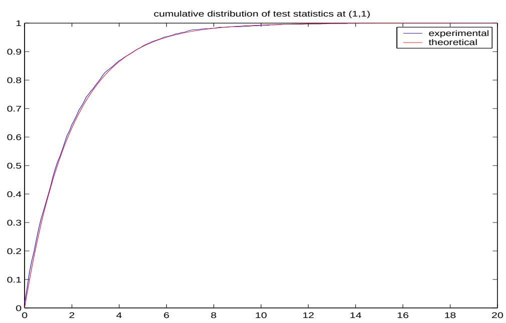

Fig. 3. Experimental and theoretical cumulative distributions of the test statistics.

4.2. Application to detecting crosstalk

In the following example, we use two music instrument sounds S_1 and S_2 which are synthesized on the computer. Figure 4 shows their waveforms (16KHz sampling rate, 3 secs, 48000 points).

Figure 5 shows the real part and the imaginary part of the scaled residual of the joint and marginal characteristic functions

$$Z_n(t, s) = \{c_n^{XY}(t, s) - c_n^X(t)c_n^Y(s)\}\sqrt{n},$$
$$(t, s) \in [-4, 4] \times [-4, 4],$$

where $X = S_1$ and $Y = S_2$ and the empirical characteristic functions are calculated from randomly chosen 5000 points of X and Y. We can see that both of the values are almost 0 around the origin, but continuously fluctuate toward the marginal region.

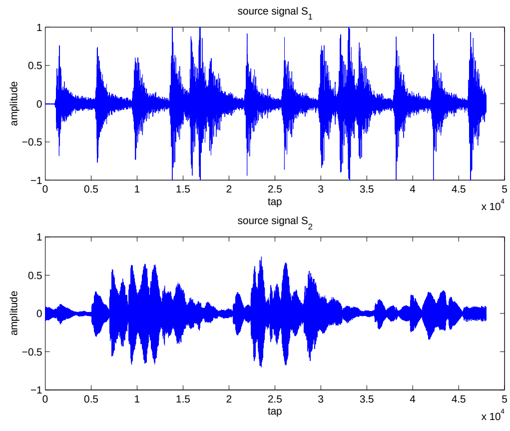

Fig. 4. Source signal. (16KHz sampling rate, 3secs waveforms)

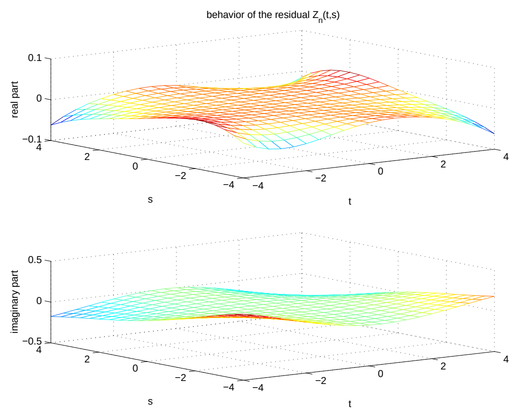

Fig. 5. The real part and the imaginary part of the residual $Z_n(t, s)$ of the joint and marginal characteristic functions.

Using these signals, we generate artificially mixed signals

$$\begin{pmatrix} \hat{S}_1 \\ \hat{S}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.04 & 1 \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}.$$

Since $\hat{S}_2$ includes S_1 with small amplitude, it simulates crosstalk in S_2. To detect the crosstalk, we generate mixed signals

$$X = \hat{S}_1,$$
$$Y = \hat{S}_2 + \alpha \hat{S}_1,$$

and calculate the statistics $T(t,s)$. In Figures 6 and 7, the test statistics and the p-values are shown, which are calculated from randomly chosen 5000 points. We plot the intensity of α in the horizontal axis and the test statistics and the *p*-values in the vertical axis, respectively, where α varies on $[-0.1, 0.1]$ and the statistics are evaluated at $(t,s) = (0.1, 0.1), (0.1, 0.2), (0.2, 0.1)$. The p-values of the test statistics are maximized at $-0.039, -0.038, -0.039$ at the applied points, respectively, hence the test statistics detect the true crosstalk ratio 0.04 with small errors.

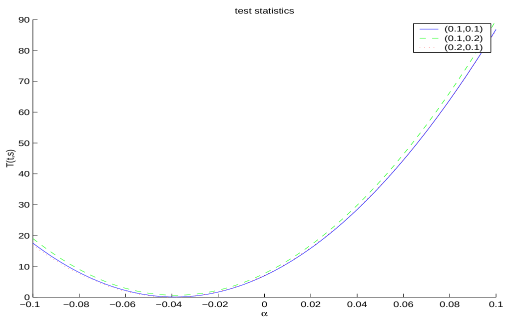

Fig. 6. Test statistics of artificially mixed signals.

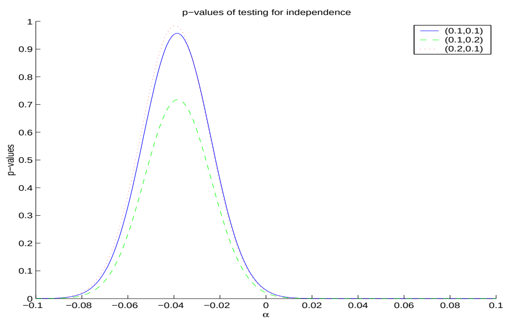

Fig. 7. *p*-values of artificially mixed signals.

5. CONCLUDING REMARKS

In this paper, we investigated the properties of the joint and marginal empirical characteristic functions, and discussed a test for independence using the characteristic function of joint and marginal empirical distributions.

With more detailed evaluation of the convergence, quantitative consideration about the relationship between the size of sample n and the evaluation point (t,s) are needed as a future work. Also, especially for practical use, it is important to consider the problem in which the observations are smeared by not negligible noises. In the Gaussian noise case, a straightforward extension can be made by using the information of the noise covariance estimated by the conventional factor analysis as discussed in Kawanabe[9],

6. REFERENCES

[1] Pierre Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.

[2] Anthony J. Bell and Terrence J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[3] Shun-ichi Amari, Andrzej Cichocki, and Harward Hua Yang, "A new learning algorightm for blind signal separation," in *Advances in Neural Information Processing Systems 8*, David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, Eds., pp. 757–763. MIT Press, Cambridge MA, 1996.

[4] Christian Jutten and Jeanny Herault, "Separation of sources, Part I," *Signal Processing*, vol. 24, no. 1, pp. 1–10, July 1991.

[5] Jean-François Cardoso and Antoine Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, December 1993.

[6] L. Féty and J. P. Van Uffelen, "New methods for signal separation," in *Proceedings of 4th International Conference on HF radio systems and techniques*, London, April 1988, IEE, pp. 226–230.

[7] Andrey Feuerverger and Roman A. Mureika, "The empirical characteristic function and its applications," *The Annals of Statistics*, vol. 5, no. 1, pp. 88–97, 1977.

[8] Kazuo Murota and Kei Takeuchi, "The studentized empirical characteristic function and its application to test for the shape of distribution," *Biometrika*, vol. 68, no. 1, pp. 55–65, 1981.

[9] Motoaki Kawanabe and Noboru Murata, "Independent component analysis in the presence of gaussian noise based on estimating functions," in *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, Helsinki, June 2000.