A BAYESIAN PROCEDURE TO RECOGNIZE INDEPENDENT SIGNALS

Chin-Jen Ku, Terrence L. Fine[*]

School of Electrical and Computer Engineering
Cornell University. Ithaca, NY. USA.
e-mails: kucj@ece.cornell.edu, tlfine@ece.cornell.edu

ABSTRACT

We propose a Bayesian test to assess the statistical dependence when only a small number of samples are available. Our procedure converts the problem of independence test to a parametric one through quantization and computes the likelihood of the observed cell counts under the independence hypothesis where the marginal cell probabilities are modeled by independent symmetric Dirichlet priors. We tested the ability of our Bayesian test in validating the solutions to the problem of blind source separation. The experimental results indicate that while the standard parametric method frequently fails to distinguish the case of independent signals from dependent ones due to the deviation of the test statistic from its desired distribution, our approach can overcome the scarcity of data samples with a proper selection of the prior parameters to achieve a significantly superior performance.

1. INTRODUCTION

Recently, many independent component analysis (ICA) algorithms have been successfully developed to solve the problem of blind source separation (BSS) in the basic case of time-invariant, instantaneous and linear mixing process. It has been observed that their performance may deteriorate when the source signals have some particular distributions or if the number of available samples is very limited [1, 2]. The later situation may arise, for instance, if the mixing matrix varies during source emission such that the basic BSS model can only remain a good approximation within a relatively small block of observed samples. Therefore, designing an independence test for small dataset offers the possibility to validate in an unsupervised fashion (*i.e.* with little or no knowledge on the sourcess and the mixing process) the estimated solutions to the BSS problem. Many nonparametric tests have been proposed and they are based on the discrepancy between the joint and the product of marginal statistical quantities [1, 3, 4, 5, 6]. They typically require a substantial amount of computation and data samples to yield reliable estimates and assume unique statistical distribution for each source, which considerably limits their application to non-iid signals. In contrast, the parametric approaches reduce the problem complexity by transforming the observed data into a table of cell counts [7, 8, 9, 10]. In [2], we examined the application of different independence tests to the problem of validating the BSS solutions. We observed that the parametric tests are numerically simpler to implement and their application may be extended to some non-iid signals. However, their use also requires a sufficient number of samples to avoid sparseness, which typically occurs when the amount of observed samples is limited or when the table size increases (*e.g.* the number of sources in presence is large). Here we present a new Bayesian test of independence to cope with small datasets. The new approach quantizes the original data into a table of cell counts and exploits the prior information embedded in the parameters that characterize the cell counts to evaluate the test statistic.

[*] Corresponding author, research of both authors partially supported by AFOSR Grant F30602-00-2-0558

2. BAYESIAN TEST STATISTIC

2.1. Formulation

The implementation of a Bayesian parametric test of independence requires the samples from all variables being quantized into a number of level sets and a prior distribution π over the parameters defining the parametric model for the joint probability of observing those quantized samples. Let us first define N as the number of signals and L as the number of observed samples. In the bivariate case (N=2), we uniformly quantize the L observed values of each variable into q levels $\{x_i\}_{i=1}^{q}$, $\{y_i\}_{i=1}^{q}$. We further define p_{ij} as the probability that a pair of sample is quantized to the cell (x_i, y_j) and N_{ij}, n_{ij} as the random variable and the observed value for the number of occurrences of data samples being discretized as (x_i, y_j). Typically, the $\{N_{ij}\}$ follow the *multinomial* distribution [9, 10] and the cell counts $\{n_{ij}\}$ may be represented by a $q \times q$ table T with q^2 cells. To build the Bayesian test statistic, a prior distribution π should be selected for the *first-order marginal probabilities* which can be easily derived from the cell probabilities $\{p_{ij}\}$:

$$p_{i+} = \sum_{j=1}^{q} p_{ij},\ p_{+j} = \sum_{i=1}^{q} p_{ij},\ i,j = 1 \ldots q. \qquad (1)$$

The new test statistic Σ is defined as the probability of observing the cell counts $\{n_{ij}\}$ under the independence hypothesis H_0:

$$\Sigma(\{n_{ij}\}) = P(\{N_{ij} = n_{ij}\} \mid H_0). \qquad (2)$$

A small Σ implies a low probability of observing the data $\{n_{ij}\}$ under H_0, thus suggesting that the signals are not statistically independent. The Bayesian test of independence consists of rejecting H_0 if $\Sigma(\{n_{ij}\}) < \tau_\gamma$, where the threshold τ_γ is determined so as to achieve the desired significance level γ. The quantity in (2) is computed by averaging over all possible distributions of the marginal cell probabilities. We choose to model the prior of both vectors of marginal cell probabilities $\mathbf{p_1} = (p_{1+}\ p_{2+}\ \ldots\ p_{q+})$, $\mathbf{p_2} = (p_{+1}\ p_{+2}\ \ldots\ p_{+q})$ by the Dirichlet distribution which may be regarded as the conjugate prior of the parameters of a multinomial distribution. The Dirichlet distribution is characterized by a vector parameter α which is commonly decomposed as the product of a positive scalar s and a probability vector $\mathbf{t}$=$(t_1 \ldots t_q)^T$ such that $0 < t_1, \ldots, t_q < 1$ and $\sum_{i=1}^{q} t_i$=1. The vector $\mathbf{t}$ and the scalar s are respectively called the *prior cell probability vector* and the *prior strength*. The prior cell probability vector $\mathbf{t}$ expresses our initial assumption about the true cell probabilities associated to the table of cell counts while the prior strength s indicates the number of (hypothetical) samples that support the claim on the choice of the vector $\mathbf{t}$. Typically, most of the objective Bayesian methods for a fixed number of data samples arising from a multinomial distribution use the *symmetric* Dirichlet priors [11] where all the entries of the vector $\mathbf{t}$ are equal to 1/q.

2.2. Test implementation

Let us define $\mathcal{S}_1, \mathcal{S}_2$ as the (q-1)-dimensional simplex domains associated to $\mathbf{p_1}$ and $\mathbf{p_2}$, respectively. Under the hypothesis of independence H_0, p_{ij} is equal to $p_{i+}p_{+j}$, hence the test statistic $\Sigma(\{n_{ij}\})$ in (2) can be written as

$$\int_{\mathcal{S}_1 \times \mathcal{S}_2} \frac{L!}{\prod_{i,j=1}^{q} n_{ij}!} \prod_{i,j=1}^{q} (p_{1i} p_{2j})^{n_{ij}} f(\mathbf{p_1}, \mathbf{p_2})\, d\mathbf{p_1} d\mathbf{p_2}. \qquad (3)$$

where f is the joint distribution for $\mathbf{p_1}$ and $\mathbf{p_2}$. By modeling $\mathbf{p_1}, \mathbf{p_2}$ as independent symmetric Dirichlet vectors with prior parameters α_1, α_2, (3) can be decoupled as the product of two integrals, each with respect to one set of marginal probabilities in virtue of the product factorization provided by the independence hypothesis. Using the definition of Dirichlet distribution, one obtains

$$\Sigma(\{n_{ij}\}) = \frac{L!}{\prod_{i,j=1}^{q} n_{ij}!} \left(\frac{\Gamma(s)}{\Gamma(s+L)}\right)^2 \times \prod_{i=1}^{q} \left(\frac{\Gamma(\alpha_{1i} + n_{+i})}{\Gamma(\alpha_{1i})}\right) \prod_{j=1}^{q} \left(\frac{\Gamma(\alpha_{2j} + n_{+j})}{\Gamma(\alpha_{2j})}\right). \qquad (4)$$

where $\Gamma(\cdot)$ is the Gamma function and α_{ij} is the jth component of the vector α_i. The variables n_{i+}, n_{+j} are the first-order marginal counts by analogy to p_{i+}, p_{+j} and they can be readily computed using (1) by substituting n_{ij} for p_{ij}. The extension of (4) to the multivariate case is straightforward. We note that the typically very small value of $\Sigma(\{n_{ij}\})$ is best evaluated through its logarithm. In addition, the computation of (4) can be greatly simplified by choosing s as an integer multiple of q. Finally, the test threshold τ_γ is estimated from the empirical distribution of $\Sigma(\{N_{ij}\})$, where the counts $\{N_{ij}\}$ are random variables that arise from a multinomial distribution with parameters given by the products of Dirichlet marginal cell probabilities. In practice, τ_γ is evaluated by repeatedly generating *iid* Dirichlet marginal prior vectors in order to compute $\{p_{ij}\}$. For each configuration of the cell probabilities, a realization of $\{N_{ij}\}$ is then performed and the test statistic computed. The threshold is obtained from the tail distribution of $\Sigma(\{N_{ij}\})$ with the appropriate value of γ.

3. RESULTS

3.1. Experimental setup

We examine the performance of the Bayesian test in the context of unsupervised validation of the BSS solutions. The model under consideration consists of a time-invariant, instantaneous and linear mixing process without noise. We recall that although this simple model does not hold in general throughout the entire signal duration, it can be considered as a fairly good approximation over a small time duration, which typically corresponds to a small amount of observed dataset. Formally, the mixing process is represented by a full rank M-by-N constant mixing matrix A ($M \geq N$) and the sources samples by an N-by-L matrix S. The observed mixtures are given by the matrix product $X = AS$. The BSS problem is solved by finding an N-by-M demixing matrix W such that the rows of the matrix $Y = WX$ form an estimate of the original data stored in S up to some row-wise scalar multiplication and permutation:

$$Y = WX = WAS = CS \qquad (5)$$

The N-by-N square matrix $C = WA$ is commonly called the *global transfer matrix*. We apply both the Bayesian test and the standard χ^2-based test using the power divergence (PD) family of test statistics [9] with parameter λ=2/3 on the quantized version of the data in Y with three levels of quantization for each signal (q=3). The degree of statistical dependence among the recovered signals is controlled by the so-called *index of minimum signal-to-interference ratio* (MSIR) [12] that measures the quality of source separation using the worst-case energy ratio between each of the recovered signal and the residual interference level from other recovered signals. Typically, an MSIR index above 30dB signifies successful separation whereas an MSIR index below 5dB entails unsatisfactory demixing results. We compare the two tests based on their rejection rate of the independence hypothesis for different combinations of the parameters γ, s, and MSIR. Three values of γ are considered: 5%, 1% and 0.1%. Ideally, at low MSIR (typically smaller than 5dB) the rejection rate should be close to 100%. As the MSIR index increases, the rejection rate should diminish and approach γ in virtue of the definition of the significance level.

3.2. Performance evaluation

We simulate different scenarios of demixing results by varying the MSIR index from 0 to 40dB. For each value of MSIR, 10,000 pairs of global transfer matrices and independent sources (C, S) are randomly and independently generated. The source samples have *iid* uniform distribution $\mathcal{U}(-0.5, 0.5)$ and each generated source sequence has zero mean. The entries of C are *iid* uniformly distributed in the interval [-1,1] and selected so that C remains invertible. Tables 1 shows the rejection rate ρ versus MSIR for both the PD test and the Bayesian test (with different values of s) in the case of six sources (N=6), q=3 and L=200. The PD-based test performs very poorly at low MSIR since ρ is always close to 0. This result is explained by the sparseness of the cell counts which causes a large deviation of the PD test statistic from its desired distribution. In contrast, our new test shows considerable improvement using the same amount of data samples; with as few as 200 samples and a properly selected prior strength s (ranging from 720 to 900), the rejection rate is almost 100% at low MSIR and close to the significance level at high MSIR.

The result from Table 1 shows that a properly chosen prior parameter can effectively compensate for the scarcity of data samples and enable the Bayesian test to validate the solutions to the BSS problem with a considerably higher accuracy. Moreover, we observe that the total prior strength s has an important effect on the test behavior. Overall, choosing a small value of s leads to a higher rejection rate of the independence hypothesis and vice versa. In the context of validating the solutions to the BSS problem, the choice of s may depend on whether the emphasis is to detect unsuccessful (dependent signals) or successful (independent signals)

Table 1. **Rejection rate (%) of the independence hypothesis versus MSIR (dB) at different significance level γ. First row: standard chisquare-based PD test. Second to sixth rows: Bayesian test with different values of the prior strength s ranging from 120 to 900 (N=6, q=3 and L=200).**

	MSIR (dB)	$\gamma = 5\%$	$\gamma = 1\%$	$\gamma = 0.1\%$
PD	0	0.08	0	0
	10	0	0	0
	20	0	0	0
	30	0	0	0
	40	0	0	0
120	0	100	99.64	97.84
	10	85.04	55.78	22.74
	20	57.14	22.12	4.54
	30	53.16	18.22	2.96
	40	52.58	18.40	3.50
300	0	99.48	97.72	91.48
	10	46.78	22.92	6.94
	20	16.58	4.82	0.72
	30	13.44	3.62	0.44
	40	13.58	3.36	0.40
600	0	98.68	95.16	87.88
	10	32.86	14.24	4.96
	20	8.68	2.28	0.18
	30	6.68	1.72	0.26
	40	7.02	1.50	0.24
720	0	98.40	94.90	82.18
	10	29.19	11.90	2.08
	20	7.96	1.66	0.12
	30	5.38	0.94	0.08
	40	6.10	1.16	0.08
900	0	98.28	94.58	84.90
	10	28.94	11.28	2.88
	20	6.04	0.92	0.06
	30	5.38	1.18	0.08
	40	5.22	0.96	0.12

demixing solutions. A smaller s leads to a more stringent test threshold, therefore a lower probability of missing unsuccessful demixing at the expense of a higher probability of rejecting suitable solutions.

4. CONCLUSION

We proposed a parametric Bayesian test of statistical independence that can perform unsupervised validation of the solutions to the basic blind source seapration problem when only a small dataset is available. The new algorithm addresses the issue that the standard approaches fail in this case due to an inaccurate assessment of the source statistical models or the parameters used by the test statistics. Our procedure first converts the observed data to a table of cell counts. It computes the likelihood of observing these cell counts assuming that the signals are independent and the marginal cell probabilities have prior *iid* Dirichlet distribution. We derived the test statistic in closed form and computed the test threshold through simulation. The experimental results indicate that provided an adequately chosen prior parameter, our new method can overcome the scarcity of data samples and significantly outperform the standard chi-square test.

5. REFERENCES

[1] J. Eriksson, A. Kankainen and V. Koivunen, “Novel Characteristic Function Based Criteria for ICA,” *Proceedings of ICA2001*, pp. 108-113, San Diego, USA, December 2001.

[2] C.-J. Ku and T.L. Fine, “Testing stochastic independence: application to blind source separation,” *IEEE Trans. on Signal Processing*, vol. 23, no. 5, 2005.

[3] N. Murata, “Properties of the empirical characteristic function and its application to testing for independence, ” *Proceedings of ICA2001*, San Diego, USA, December 2001.

[4] M. Rosenblatt and B.E. Wahlen, “A nonparametric measure of independence under a hypothesis of independence components,” *Statistics and Probability Letters*, Vol. 15, pp. 245-252, 1992.

[5] G.R. Heer, “Testing independence in high dimensions,” *Statistics and Probability Letters*, Vol. 12, pp. 73-81, 1991.

[6] B. Fernández Fernández and J.M. González-Barrios, “Multidimensional dependency measures,” *Journal of Multivariate Analysis*, vol. 89, pp. 351-370, 2004.

[7] J. Baglivo, D. Olivier and M. Pagano, “Methods for exact goodness-of-fit tests,” *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 464-469, 1992.

[8] J. Baglivo, D. Olivier and M. Pagano, “Methods for the analysis of contingency tables with large and small cell counts,” *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1006-1013, 1988.

[9] T.R.C. Read and N.A.C. Cressie, *Goodness-of-fit statistics for discrete multivariate analysis*, Springer-Verlag, 1988.

[10] Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, *Discrete multivariate analysis: theory and practice*, MIT Press, 1975.

[11] J.-M. Bernard, ”Analysis of local or asymmetric dependencies in contingency tables using the imprecise Dirichlet model,” *Proceedings of ISIPTA 2003*, Lugano, Switzerland, July 2003.

[12] C.-J. Ku and T. L. Fine, “A new ICA algorithm for blind source separation,” *Proceedings of ICASSP 2003*, Hong Kong, April 2003.