Bayesian source separation of functional sources

Daniel B. Rowe

Department of Biophysics

Medical College of Wisconsin

8701 Watertown Plank Rd.

Milwaukee, WI 53226

U. S. A.

Abstract

This paper incorporates available prior knowledge of the source waveforms into the Bayesian approach to blind source separation. The source separation model is described, prior distributions are introduced to quantify available prior knowledge regarding the model parameters, the posterior distribution for the model parameters is formed, and parameter estimation is detailed. Finally, the methods discussed are applied to an example.

1. Introduction and model

The problem addressed by source separation is that of separating unobservable or latent source signals when mixed signals are observed. In other words, to take a set of observed mixed signal vectors and unmix or separate them into a set of true unobservable source signal vectors. This paper adopts a Bayesian statistical approach and the linear synthesis model.

For motivation and illustration of the source separation model, the context of the "cocktail party problem" is adopted. At a cocktail party, there are p microphones that record or observe m partygoers or speakers at n time increments. The observed conversations consist of mixtures of true conversations. In other words, p-dimensional mixed signal vectors $x_i = (x_{i1}, \ldots, x_{ip})'$ are observed and the goal is to separate these observed signal vectors into m-dimensional true underlying source signal vectors, $s_i = (s_{i1}, \ldots, s_{im})'$ where $i = 1, \ldots, n$.

Journal of Interdisciplinary Mathematics
Vol. 6 (2003), No. 2, pp. 129–138

The linear synthesis source separation model is

$$\underset{(p \times 1)}{(x_i \mid \Lambda, s_i)} = \underset{(p \times m)}{\Lambda} \; \underset{(m \times 1)}{s_i} + \underset{(p \times 1)}{\epsilon_i}, \tag{1.1}$$

where it has been assumed that the observed signals have had their mean and liner trend subtracted off if they exist, further

$\Lambda =$ a $p \times m$ matrix of unobserved mixing constants, $\Lambda = (\lambda'_1, \ldots, \lambda'_p)'$;

$s_i =$ the i^{th} m-dimensional unobservable source vector, $s_i = (s_{i1}, \ldots, s_{1m})'$;

$\epsilon_i =$ the p-dimensional vector of errors or noise terms of the i^{th} observed signal vector $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})'$.

2. Likelihood, priors and posterior

It is specified that the errors of the observations are independent over time and motivated by the central limit theorem, normally distributed random vectors with zero and covariance matrix Ψ, thus the likehood of a given observation vector is

$$p(x_i \mid \Lambda, s_i, \Psi) \propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i - \Lambda s_i)'\Psi^{-1}(x_i - \Lambda s_i)}. \tag{2.1}$$

The joint likelihood of the observations is

$$p(x_1, \ldots, x_n | \Lambda, s_1, \ldots, s_n, \Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \Lambda s_i)'\Psi^{-1}(x_i - \Lambda s_i)}. \tag{2.2}$$

Analogous to regression, the source separation model can be written in terms of matrices as

$$\underset{(n \times p)}{(X \mid \Lambda, S)} = \underset{(n \times m)}{S} \; \underset{(m \times p)}{\Lambda'} + \underset{(n \times p)}{E}, \tag{2.3}$$

with likelihood

$$p(X \mid \Lambda, S, \Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - S\Lambda')'(X - S\Lambda')} \tag{2.4}$$

where $X' = (x_1, \ldots, x_n)$, $S' = (s_1, \ldots, s_n)$, and $E' = (\epsilon_1, \ldots, \epsilon_n)$. The time series of observations for the j^{th} microphone is the j^{th} column of X and the time series of unobservables for the k^{th} source is the k^{th} column of S.

Available knowledge regarding the parameter values is incorporated into the inferences in the form of prior distributions.

The source vectors s_i are specified to be normally distributed with mean s_{i0} and covariance matrix R represented by

$$(s_i \mid R) \sim N(s_{i0}, R) \tag{2.5}$$

and distribution function given by

$$p(s_i \mid R) \propto |R|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(s_i - s_0)' R^{-1} (s_i - s_0)} \,. \tag{2.6}$$

The joint prior probability distribution function for all the unobserved source signal vectors is

$$p(s_1, \ldots, s_n \mid R) \propto |R|^{-\frac{n}{2}} \, e^{-\frac{1}{2}\sum_{i=1}^{n}(s_i - s_{i0})' R^{-1} (s_i - s_{i0})} \,. \tag{2.7}$$

The prior distribution for the source vectors can also be rewritten as a matrix normal distribution

$$p(S \mid R) \propto |R|^{-\frac{n}{2}} \, e^{-\frac{1}{2} tr (S - S_0) R^{-1} (S - S_0)'} \tag{2.8}$$

where $S_0' = (s_{10}, \ldots, s_{n0})$.

Again, the objective is to unmix the sources by estimating S and to obtain knowledge about the mixing process by estimating Λ and Ψ.

Regarding the other parameters, information is incorporated as an inverted Wishart distribution for the covariance of the source vectors with η degrees of freedom and scale matrix V, an inverted Wishart distribution for the covariance of the observed vectors with ν degrees of freedom and scale matrix B, and a normal distribution for the mixing matrix with mean Λ_0 and covariance $\Psi \otimes H^{-1}$. The inverted Wishart distribution is the multivariate generalization of the inverted gamma distribution which is used as a prior distribution for variances.

All together, the prior distribution are

$$p(S \mid R) \;\propto\; |R|^{-\frac{n}{2}} \, e^{-\frac{1}{2}(S - S_0) R^{-1} (S - S_0)'} \,, \tag{2.9}$$

$$p(R) \;\propto\; |R|^{-\frac{\eta}{2}} \, e^{-\frac{1}{2} tr R^{-1} V} \,, \tag{2.10}$$

$$p(\Psi) \;\propto\; |\Psi|^{-\frac{\nu}{2}} \, e^{-\frac{1}{2} tr \Psi^{-1} B} \,, \tag{2.11}$$

$$p(\Lambda \mid \Psi) \;\propto\; |\Psi|^{-\frac{m}{2}} \, e^{-\frac{1}{2} tr \Psi^{-1} (\Lambda - \Lambda_0) H (\Lambda - \Lambda_0)'} \,, \tag{2.12}$$

where $S_0, \eta, V, \nu, B, \Lambda_0$ and H are hyperparameters to be assessed.

Available information as to a functional form for the sources can now be incorporated. For example, if the k^{th} source varied sinusoidally then this is incorporated as

$$s_{ik0} = \sin(a_k i + b_k) \tag{2.13}$$

or as a square wave

$$s_{ik0} = \begin{cases} 1, & \text{for } a_{kl} < i < p_{kl} \\ -1, & \text{otherwise} \end{cases}$$

for $l = 1, \ldots, L$.

Upon using Bayes' rule the posterior distribution for the unknown parameters is

$$p(\Lambda, S, \Psi, R \mid X) \propto |\Psi|^{-\frac{(n+m+\nu)}{2}} |R|^{-\frac{n+\eta}{2}} e^{-\frac{1}{2}tr\Psi^{-1}U} \cdot e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \tag{2.14}$$

where

$$U \equiv (X - S\Lambda')'(X - S\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B\,.$$

This posterior distribution must now be evaluated in order to obtain parameter estimates of the mixing matrix, the sources, the errors of the sources, and the errors of observation.

3. Estimation

With the above posterior distribution, it is not possible to obtain marginal distributions and thus marginal estimates for any of the parameters in an analytic closed form. For this reason, the maximization technique iterated conditional modes (ICM) is used (Lindley and Smith, 1972) to obtain maximum a posteriori estimates. For the ICM estimation procedure, the posterior conditional distributions are required.

Form the joint posterior distribution we can obtain the posterior conditional distributions. The conditional posterior distributions for the mixing matrix is

$$\begin{aligned} p(\Lambda \mid S, \Psi, R, X) &\propto p(\Lambda, S, \Psi, R)p(X \mid \Lambda, S, \Psi) \\ &\propto p(\Psi)p(\Lambda \mid \Psi)p(R)p(S \mid R)p(X \mid \Lambda, S, \Psi) \\ &\propto p(\Lambda \mid \Psi)p(X \mid \Lambda, S, \Psi) \\ &\propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \end{aligned}$$

$$\cdot \quad |\Psi|^{-\frac{n}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}(X-S\Lambda')'(X-S\Lambda')}$$
$$\propto \; e^{-\frac{1}{2}tr\Psi^{-1}[(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'+(X-S\Lambda')'(X-S\Lambda')]}$$
$$\propto \; e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\tilde{\Lambda})(H+S'S)(\Lambda-\tilde{\Lambda})'} \tag{3.1}$$

where the posterior conditional mean and mode is given by

$$\tilde{\Lambda} = [X'S + \Lambda_0 H](H + S'S)^{-1} . \tag{3.2}$$

The conditional distribution for the mixing matrix given the other parameters and the data is normally distributed.

The conditional posterior distribution of the observation error matrix is

$$\begin{aligned} p(\Psi \mid \Lambda, S, R, X) &\propto p(\Lambda, S, \Psi, R)p(X \mid \Lambda, S, \Psi) \\ &\propto p(\Psi)p(\Lambda \mid \Psi)p(R)p(S \mid R)p(X \mid \Lambda, S, \Psi) \\ &\propto p(\Psi)p(\Lambda \mid \Psi)p(X \mid \Lambda, S, \Psi) \\ &\propto |\Psi|^{-\frac{\nu}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}B}|\Psi|^{-\frac{m}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \\ &\cdot \quad |\Psi|^{-\frac{n}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}(X-S\Lambda')'(X-S\Lambda')} \\ &\propto |\Psi|^{-\frac{(n+m+\nu)}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}U} \end{aligned} \tag{3.3}$$

where

$$U = (X - S\Lambda')'(X - S\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B \tag{3.4}$$

with a mode given by

$$\tilde{\Psi} = \frac{U}{n + m + \nu} . \tag{3.5}$$

The conditional distribution of the observation error covariance matrix given the other parameters and the data is an inverted Wishart.

The conditional posterior distribution for the sources is

$$\begin{aligned} p(S \mid \Lambda, \Psi, R, X) &\propto p(\Lambda, S, \Psi, R)p(X \mid \Lambda, S, \Psi) \\ &\propto p(\Psi)p(\Lambda \mid \Psi)p(R)p(S \mid R)p(X \mid \Psi, S, \Lambda) \\ &\propto p(S \mid R)p(X \mid \Lambda, S, \Psi) \\ &\propto |\Psi|^{-\frac{n}{2}} \, e^{-\frac{1}{2}tr\Psi^{-1}(X-S\Lambda')'(X-S\Lambda')} \end{aligned}$$

$$
\begin{aligned}
&\propto |\Psi|^{-\frac{n}{2}} |R|^{-\frac{n}{2}} \\
&\quad \times e^{-\frac{1}{2}tr[(S-S_0)R^{-1}(S-S_0)'+(X-S\Lambda')\Psi^{-1}(X-S\Lambda')']} \\
&\propto e^{-\frac{1}{2}tr(S-\widetilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\widetilde{S})'}
\end{aligned} \tag{3.6}
$$

where the posterior conditional mean and mode is given by

$$
\widetilde{S} = (X\Psi^{-1}\Lambda + S_0 R^{-1})(R^{-1} + \Lambda'\Psi^{-1}\Lambda)^{-1} . \tag{3.7}
$$

The conditional posterior distribution for the sources given the other parameters and the data is normally distributed.

The conditional posterior distribution for the source covariance matrix is

$$
\begin{aligned}
p(R \mid \Lambda, S, \Psi, X) &\propto p(\Psi, \Lambda, S, R) p(X \mid \Lambda, S, \Psi) \\
&\propto p(\Psi) p(\Lambda \mid \Psi) p(R) p(S \mid R) p(X \mid \Psi, S, \Lambda) \\
&\propto p(R) p(S \mid R) \\
&\propto e^{-\frac{1}{2}trR^{-1}V} |R|^{-\frac{\eta}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} |R|^{-\frac{n}{2}} \\
&\propto |R|^{-\frac{(n+\eta)}{2}} e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]}
\end{aligned} \tag{3.8}
$$

with the posterior conditional mode given by

$$
\widetilde{R} = \frac{(S - S_0)'(S - S_0) + V}{n + \eta} . \tag{3.9}
$$

The conditional posterior distribution for the source covariance matrix given the other parameters and the data is inverted Wishart distributed.

The ICM estimation procedure consists of starting with initial values for S, and R, say $\widetilde{S}_{(0)}$, $\widetilde{R}_{(0)}$ and cycling through

$$
\begin{aligned}
\widetilde{\Lambda}_{(l+1)} &\equiv (X'\widetilde{S}_{(l)} + \Lambda_0 H)(H + \widetilde{S}'_{(l)}\widetilde{S}_{(l)})^{-1} \\
\widetilde{\Psi}_{(l+1)} &\equiv \frac{\begin{pmatrix} (X - \widetilde{S}_{(l)}\widetilde{\Lambda}'_{(l+1)})'(X - \widetilde{S}_{(l)}\widetilde{\Lambda}'_{(l+1)}) \\ +(\widetilde{\Lambda}_{(l+1)} - \Lambda_0)H(\widetilde{\Lambda}_{(l+1)} - \Lambda_0)' + B \end{pmatrix}}{n + m + \nu} \\
\widetilde{S}_{(l+1)} &\equiv (X\widetilde{\Psi}^{-1}_{(l+1)}\widetilde{\Lambda}_{(l+1)} + S_0\widetilde{R}^{-1}_{(l)})(\widetilde{R}^{-1}_{(l)} + \widetilde{\Lambda}'_{(l+1)}\widetilde{\Psi}^{-1}_{(l+1)}\widetilde{\Lambda}_{(l+1)})^{-1} \\
\widetilde{R}_{(l+1)} &\equiv \frac{(\widetilde{S}_{(l+1)} - S_0)'(\widetilde{S}_{(l+1)} - S_0) + V}{n + \eta}
\end{aligned}
$$

until convergence is reached. The converged values $(\tilde{S}, \tilde{\Lambda}, \tilde{\Psi}, \tilde{R})$ are joint posterior modal or maximum a posteriori estimators of the parameters.

4. Example

For an example, a simulation was carried out in which there were $m = 3$ sources and $n = 128$ observations of dimension $p = 1$ with known true parameter values $\Lambda_T = (20, 50, 10)$,

$$R_T = \begin{pmatrix} 0.5 & 0.25 & 0 \\ 0.25 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

and $\Psi_T = 5$. The observations were formed by generating a random s_i from $N(s_{i0}, R_T)$, premultiplying it by Λ_T, and adding an error term generated randomly form $N(0, \Psi_T)$. The true sources were sine waves of unit amplitude and frequencies $1/64$, $1/72$, and $1/20$, respectively.

The hyperparameters were assessed according to the methods in the appendix to be $\nu = 46$, $b_0 = 420$, $h_0 = 1/2$, $\Lambda_0 = (25, 45, 5)$, $\eta = 20$, and $v_0 = 60$. Further, prior mean for the sources were assessed to be square, sine, and sine waves with the same frequencies and unit amplitude.

Upon applying the ICM estimation procedure, the posterior parameter estimates were found to be

$$\begin{aligned} \tilde{\Psi} &= 2.4797 \\ \tilde{R} &= \begin{pmatrix} 0.5064 & 0.2066 & 0.0248 \\ 0.2066 & 0.8282 & 0.0508 \\ 0.0248 & 0.0508 & 0.4115 \end{pmatrix} \\ \tilde{\Lambda} &= (24.9371,\ 51.0167,\ 6.1254), \end{aligned}$$

and the true sources, mixed true sources, noisy sources, and mixed noisy sources are displayed in Figure 1 ; while the noisy sources, mixed noisy sources, prior sources, and unmixed sources are displayed in Figure 2.

5. Conclusion

It has been shown that available information as to a functional form for the source waveforms can readily be incorporated into the Bayesian source separation model. In an example, it has been shown that mixed observations can be separated out into their constituent components. Bayesian separation of sources appears to be very promising in solving the real “cocktail party problem”.

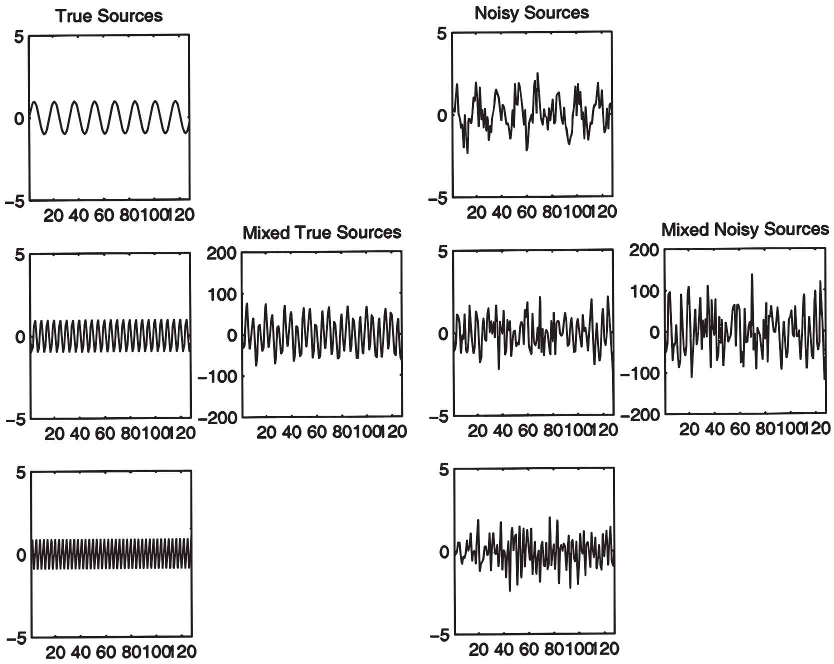

Figure 1
True sources, mixed true sources, noisy sources, and mixed noisy sources

Appendix

A *Hyperparameter assessment*

The hyperparameters of the prior distributions are assessed in this appendix.

$\boldsymbol{\Psi} : \boldsymbol{\nu}, \boldsymbol{B}.$ For simplicity of assessment, specify that $B = b_0 I_p$. With B diagonal, the mean and variance of any diagonal element of Ψ, Ψ_{jj} are

$$\begin{aligned} E(\Psi_{jj}) &= \frac{b_0}{\nu - 2p - 2} \\ \text{var}(\Psi_{jj}) &= \frac{2b_0^2}{(\nu - 2p - 2)^2(\nu - 2p - 4)} \,. \end{aligned} \tag{A.1}$$

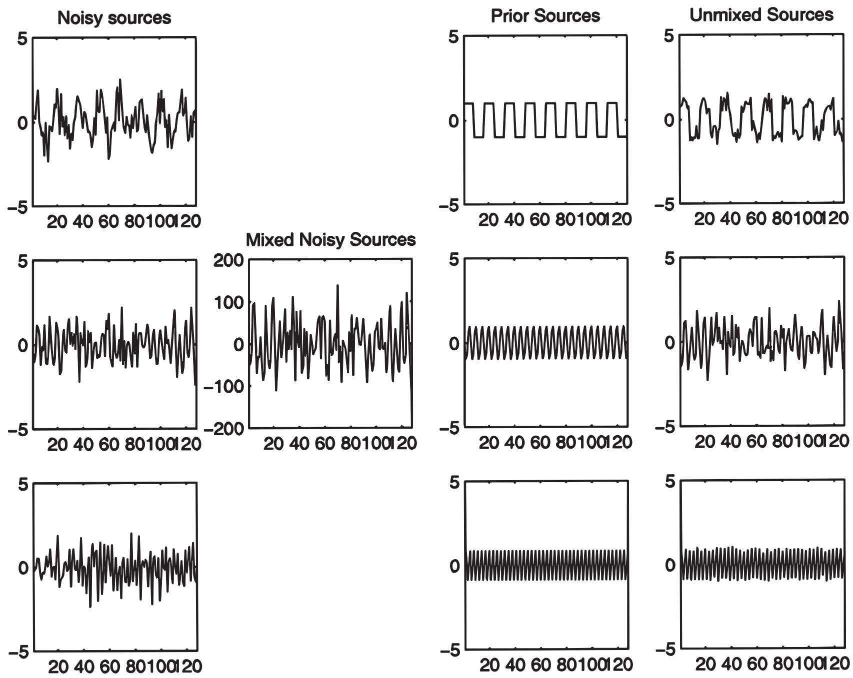

Figure 2
Noisy sources, mixed noisy sources, prior sources, and unmixed sources

Solving for ν in the above system of equations,

$$\nu = \frac{[E(\Psi_{jj})]^2}{2[\mathrm{var}(\Psi_{jj})]} + 2p + 4\,. \tag{A.2}$$

The unknowns are $E(\Psi_{jj})$ and $\mathrm{var}(\Psi_{jj})$. Prior values for the mean and variance are to be elicited from the substantive field expert. In the example, the values assessed were assumed to be $E(\Psi_{jj}) = 10$ and $\mathrm{var}(\Psi_{jj}) = 5$.

$\boldsymbol{R : \eta, V}$**.** Similarly, assessment is simplified by specifying that $V = v_0 I_m$ and thus

$$\eta = \frac{[E(R_{kk})]^2}{2[\mathrm{var}(R_{kk})]} + 2m + 4\,. \tag{A.3}$$

In the example, the values assessed were assumed to be $E(R_{kk}) = 5$ and $\mathrm{var}(R_{kk}) = 5$.

$\boldsymbol{\Lambda} : \boldsymbol{\Lambda_0}, \boldsymbol{H} = \boldsymbol{h_0 I_m}$. The hyperparameter h_0 and the prior mean for the mixing matrix Λ_0 must be subjectively elicited from the substantive field expert or estimated from training data using another technique. In the example, it was assumed that a substantive field expert provided them as $h_0 = 1/2$ and $\Lambda_0 = (25, 45, 5)$.

References

[1] K. Knuth, Bayesian source separation and localization, in *SPIE'98 Proceedings : Bayesian Inference for Inverse Problems*, A. Mohammad-Djafari (Ed.), San Diego, CA, pp. 147-158, July 1998.

[2] K. Knuth, A Bayesian approach to source separation, in *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation : ICA'99*, C. Jutten, J.-F. Cardoso and P. Loubaton (Eds.), Aussios, France, pp. 283-288, 1999.

[3] Samuel Kotz and Norman Johnson (Eds.), *Encyclopedia of Statistical Science*, Vol. 5, John Wiley & Sons, Inc., New York, 1985, pp. 326-333.

[4] D. V. Lindley and A. F. M. Smith, Bayes estimates for the linear model, *Journal of the Royal Statistical Society B*, Vol. 34 (1), 1972.

[5] D. J. C. MacKay, Maximum likelihood and covariant algorithms for independent component analysis, Technical report, University of Cambridge, Cavindish Laboratory, Cambridge, U.K., 1996.

[6] A. Mohammad-Djafari, A Bayesian estimation method for detection, localisation and estimation of superposed sources in remote sensing, in *SPIE 97 annual meeting* (San Diego, CA, U. S. A., July 27-Aug. 1, 1997), 1997.

[7] A. Mohammad-Djafari, A Bayesian approach to source separation, in *Proceedings of The Nineteenth International Conference on Maximum Entropy and Bayesian Methods* (August 2-6, 1999, Boise, ID.), 1999.

[8] Anthony O'Hagen, *Kendall's Advanced Theory of Statistics, Vol. 2B, Byesian Inference*, John Wiley & Sons Inc., New York, 1994.

[9] S. James Press, *Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference*, Robert E. Krieger Publishing Company, Malabar, Florida, 1982.

[10] S. James Press, *Bayesian Statistics : Principles, Models and Applications*, John Wiley & Sons, New York, 1989.

[11] D. B. Rowe, A Bayesian approach to blind source separation, *Journal of Interdisciplinary Mathematics*, Vol. 5 (1) (2002), pp. 49-76.

Received September, 2000