

Combinatorial Mixtures of Multiparameter Distributions: an Application to Microarray Data ¹

Valeria Edefonti

Istituto di Statistica Medica e Biometria "Giulio A. Maccacaro",
Università degli Studi di Milano, Milano, ITALY, e-mail: valeria.edefonti@unimi.it

Giovanni Parmigiani

The Sidney Kimmel Comprehensive Cancer Center,
Johns Hopkins University, Baltimore, MD, U.S.A., e-mail: gp@jhu.edu

Abstract: The term *combinatorial mixtures* refers to a flexible class of models for inference on mixture distributions whose components have multidimensional parameters. The approach is to allow each element of the component-specific parameter vector to be shared by a subset of other components. This allows for mixtures that range from very flexible to very parsimonious, and unifies the inference on component-specific parameters with that on the number of components. We develop Bayesian inference and computation approaches for this class of distributions, and illustrate them in an application based on the normal model. This work was originally motivated by the analysis of cancer subtypes: in terms of biological measures of interest, subtypes may be characterized by differences in location, scale, correlations or any of the combinations. We illustrate our approach in a simplified setting, using data on molecular subtypes of lung cancer.

Keywords: Bayesian inference, Markov Chain Monte Carlo, clustering.

1. Introduction

One of the important remaining challenges of mixture modeling is to develop approaches that achieve a practical compromise between flexibility and parsimony, especially for mixtures whose component distributions are themselves characterized by multiple parameters. In this setting, one has the option of allowing each component to have its own parameter vector, or to share a subset of the vector elements across components. For example, in the context of normal mixtures, it is common to assume either component-specific locations and variances, or a common variance, or a common mean. These three choices are extremes of a richer and useful set of patterns in which one shares some of the parameters in some of the components. Here we develop this idea formally.

2. Combinatorial Mixtures

With the term *combinatorial mixtures* we refer to a general class of mixture models in which elements of the parameter vector can be shared across any subset of the components, and positive mass is put on every possible combination of sharing patterns. We defined this class formally for any Bayesian multivariate mixture model, and detailed it

¹The present paper is financially supported by PRIN 2004, PRIN 2005 and NSF grant DMS034211.

considering the case of normal mixtures. In the following, we present the bivariate model specification.

Data $\mathbf{x}^n = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ are assumed to be independent observations from a mixture density with at most K^* bivariate normal components, $N_2(\mathbf{x}_i | \boldsymbol{\mu}_k, \Sigma_k)$. To separate standard deviations and correlations, we adopt a direct decomposition of the variance-covariance matrices, Σ_k , (Barnard *et al.* (2000)) and propose choices for the priors in terms of standard deviations, $\mathbf{S}_k$, and correlations, $\mathbf{R}_k$. In the case $K^* = 3$, after suitable independence assumptions, we can specify independent priors for $(\boldsymbol{\mu}, \mathbf{S}, \mathbf{R})$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{K^*})$, $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_{K^*})$, $\mathbf{R} = (\mathbf{R}_1, \dots, \mathbf{R}_{K^*})$, having the structure here introduced for $\boldsymbol{\mu}$:

$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3 : \boldsymbol{\mu}^j = (\mu_{1j}, \mu_{2j}, \mu_{3j}) | \gamma_j^m, \eta^2 \text{ i.i.d. } \sim$

$$\left\{ \begin{array}{ll} N(\mu_{1j} | 0, \eta^2) & \text{if } \gamma_j^m = 000 \\ N(\mu_{1j} | 0, \eta^2) N(\mu_{2j} | 0, \eta^2) & \text{if } \gamma_j^m = 110 \\ N(\mu_{1j} | 0, \eta^2) N(\mu_{2j} | 0, \eta^2) & \text{if } \gamma_j^m = 101 \\ N(\mu_{1j} | 0, \eta^2) N(\mu_{3j} | 0, \eta^2) & \text{if } \gamma_j^m = 011 \\ N(\mu_{1j} | 0, \eta^2) N(\mu_{2j} | 0, \eta^2) N(\mu_{3j} | 0, \eta^2) & \text{if } \gamma_j^m = 111 \end{array} \right. , \quad j = 1, 2,$$

$$\gamma_j^m : \gamma_j^m | \boldsymbol{\pi}^m \text{ i.i.d. } \sim \text{Multi}(1, \boldsymbol{\pi}^m), \quad \boldsymbol{\pi}^m = \begin{bmatrix} P(\mu_{1j} = \mu_{2j} = \mu_{3j}) \\ P(\mu_{1j} \neq \mu_{2j} = \mu_{3j}) \\ P(\mu_{1j} = \mu_{3j} \neq \mu_{2j}) \\ P(\mu_{1j} = \mu_{2j} \neq \mu_{3j}) \\ P(\mu_{1j} \neq \mu_{2j} \neq \mu_{3j}) \end{bmatrix}, \quad j = 1, 2,$$

where γ_j^m is the random vector that contains the sharing patterns for $\boldsymbol{\mu}^1, \boldsymbol{\mu}^2$, $\text{Multi}(\cdot | N, \mathbf{p})$ indicates a Multinomial distribution with parameters N , and $\mathbf{p}$ and all the hyperparameters are known. Similarly, for the standard deviations we choose independent log normal priors having the same structure, and, for each correlation coefficient r_k in $\mathbf{R}_k$, uniform distributions on $(-1, 1)$ building the described prior structure. Finally: $\mathbf{w} \sim \text{Dir}(\mathbf{a}_0)$, $\mathbf{a}_0 = (a_{1,0}, a_{2,0}, a_{3,0})$.

Meaningful parametric inference in mixture models requires to tackle the label switching problem (Marin *et al.* (2005)), which is the invariance of the likelihood under relabeling of the mixture components. As our main goals here are closer to clustering than to pure parameter estimation, we follow O'Hagan (1997) which proposed to summarize inferences by looking at pairs of observations and counting how often they are assigned to the same mixture component. We implement a graphical representation of the corresponding relative frequencies matrix, named the "O'Hagan" matrix from now on, to visualize clusters in the data.

Our solution does not impose constraints on either single parameters or combinations of them. If no constraints are put on the parameters, and combinatorial mixtures are assumed, apparently different combinations of γ s for means and corresponding standard deviations may actually correspond to the same mixture model for the data. Equivalent combinations are highlighted with the same color. Corresponding prior (posterior) probabilities are summed when reporting results.

Combinatorial mixtures allow for greater generality and flexibility in comparison with traditional approaches to mixture modeling, while still allowing to assign mass to models that are more parsimonious than the general mixture case in which no sharing takes place. Therefore, they are mainly suitable for those situations where many different scenarios may occur; in genomics applications, for example, this is especially critical in interpreting cancer clusters, as those may arise from changes in location, scale or correlations, or

any of the combinations.

Combinatorial mixtures apply, as a special case, in the supervised context as well. Prior independence between parameters or vague priors are not essential to combinatorial mixtures.

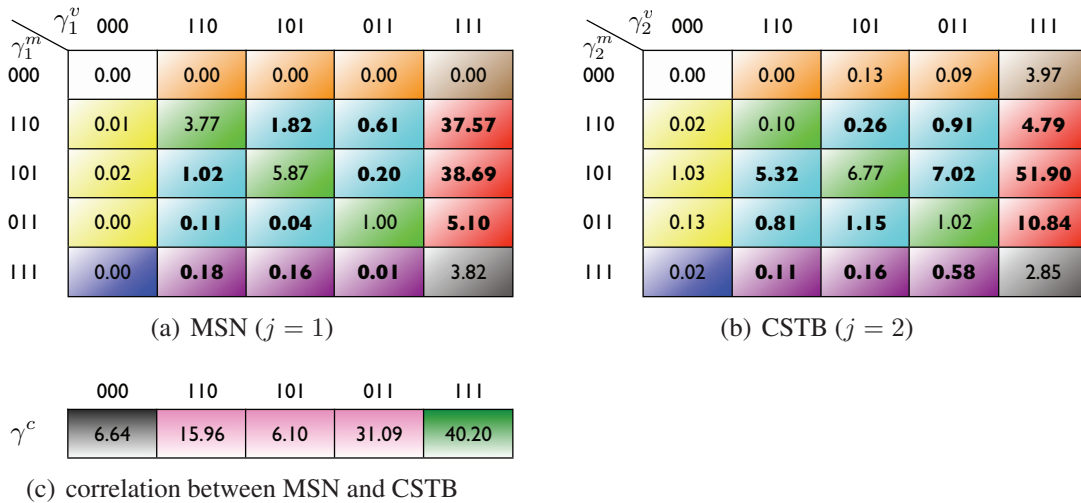
Bayesian inference for mixture models may be performed using Markov Chain Monte Carlo (MCMC) methods, generating iteratively the parameters and the missing data. We implement a suitable MCMC for the described model, updating jointly the parameters and their sharing pattern indicators. For cases, like our, where the target distribution is a mixture of mutually singular distributions, see Gottardo and Raftery (2008).

3. Gene Expression in Lung Cancer

Next, the normal mixture model of Section 2. is applied to two genes, named MSN and CSTB, coming from data on molecular classification of lung cancer (Garber *et al.* (2001)). A preliminary two-dimensional kernel density estimation plot of the bivariate distribution of these genes suggests the existence of two or more groups.

We consider Bayesian estimation in the case where we do not have strong prior information on the parameters. We run one chain for our data, doing several simulations. We report results from one of these chains, corresponding to a total of 40,000 iterations with a burn-in of 2,000 iterations. We choose the starting values for each run according to a preliminary k-means clustering algorithm. The number of groups is specified looking at some exploratory plots.

Figure 1: Joint posterior probabilities (%) of (γ_j^m, γ_j^v) , $j = 1, 2$, and posterior probability (%) of γ^c , as obtained by one run of the simulation. For correct interpretation, one needs to sum frequencies in cells with the same color. The frequencies of those cases that are gained using combinatorial mixtures, in comparison with traditional approaches, are emphasized using boldface.



The estimated joint posterior probabilities (%) of (γ_j^m, γ_j^v) , $j = 1, 2$, and the estimated

posterior probability (%) of γ^c are shown in Figure 1. The MSN marginal chain spends more than 80% of the iterations on either (110, 111) or (101, 111) or (001, 111) (red cells). The second most probable combination is: (110, 110) or (101, 101) or (011, 011) (green cells). A posteriori a mixture model with either two or three components with different means and standard deviations seems suitable for MSN. The CSTB marginal chain spends more than 65% of the iterations on either (110, 111) or (101, 111) or (001, 111) (red cells) too. The second most probable combination happens on the six light-blue cases: $(\gamma^m, \gamma^v) = (110, 101)$, $(\gamma^m, \gamma^v) = (110, 011)$ or $(\gamma^m, \gamma^v) = (101, 110)$, $(\gamma^m, \gamma^v) = (101, 011)$ or $(\gamma^m, \gamma^v) = (011, 110)$, $(\gamma^m, \gamma^v) = (011, 101)$. A posteriori a mixture model with three components with different means and standard deviations seems suitable for CSTB.

In both cases, there is not so much evidence in favor of those cases allowed by traditional mixture models. The chain spends more than 80% of the iterations on the 12 combinations of (γ_1^m, γ_1^v) and (γ_2^m, γ_2^v) that would not be possible to select using traditional approaches to mixture models. The majority of this 80% is due to the most probable combinations (red and light-blue cells). Combinatorial mixtures might end up either in extra flexibility in modeling standard deviations or in a more parsimonious solution, depending if one assumes the traditional two-component or the three-component mixture for comparison. Finally, the chain spends almost all of the iterations either on the intermediate cases of two different correlations (pink cells) or on the case of three different correlations (dark-green cells). Traditional mixture models where a decomposition of the variance-covariance matrix is assumed allow to fit all the five cases listed in Table 2(c), depending on how many different components are assumed; so, no cases are emphasized using boldface. The adoption of a decomposition allows to share one or both the standard deviations, without necessarily sharing the correlations (or viceversa), and represents an easy way to guarantee a freer modeling of the variance-covariance structure. However, all the standard deviations (correlations) should be either equal or different across components. Combinatorial mixtures allow to model each variance/correlation in a freer way, at the expense of dealing with extra complexity in model specification.

References

- Barnard J., McCulloch R.E. and Meng X.L. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage, *Statistica Sinica*, 10, 1281–1311.
- Garber M.E., Troyanskaya O.G. and et al. S.K. (2001) Diversity of gene expression in adenocarcinoma of the lung, *Proc. National Academy of Science USA*, 98, 24, 13784–13789.
- Gottardo R. and Raftery A.E. (2008) Markov chain monte carlo with mixtures of singular distributions, *Journal of Computational and Graphical Statistics*, 17, 4, 917–943.
- Marin J.M., Mengersen K. and Robert C.P. (2005) Bayesian modelling and inference on mixtures of distributions, in: *Handbook of Statistics*, Dey D. and Rao C.R., eds., Elsevier-Sciences, 459–507.
- O’Hagan A. (1997) Contribution to the discussion of richardson and green (1997), on bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society, Series B*, 59, 772.