



On the consistency of Bayes estimates for the infinite continuous mixture of Dirichlet distributions

Seloua Boukabour*, Afif Masmoudi

Probability and Statistics Laboratory, Faculty of Sciences, Sfax University, Sfax BP 1171, Tunisia

Abstract

In this paper, we introduced the infinite continuous mixture of Dirichlet distributions as a generalization of the infinite mixture of Dirichlet ones, in order to avoid the limitation of choosing the a priori sample size for the expectation *a posteriori* estimator. Monte-Carlo sampling was used in order to obtain the *posterior* distributions mixture, since this mixture is difficult to get analytically. A new parametrization of this proposed distribution was achieved. Then, we suggested a mixture expectation *a posteriori* estimator of the unknown parameters. The proposed estimator solves the problem of how to construct a Bayesian estimation of proportions without specifying particular parameters and sample size of the prior knowledge. Some asymptotic properties of this estimator were derived, specifically, its bias and variance. The consistency and asymptotic normality of the estimator were also established when the sample size tends to infinity and its credible interval was determined. The performance of the proposed estimator was illustrated theoretically and by means of a simulation study. Ultimately, a comparative simulation study between the learned estimates, the proposed mixture expectation *a posteriori*, standard Bayesian estimator, maximum likelihood and Jeffreys estimator, was established. According to this simulation, we were able to conclude that the prior infinite mixture of Dirichlet distributions offers higher accuracy and flexibility for modeling and learning data.

Mathematics Subject Classification (2020). 62F12, 62F15, 97K60.

Keywords. Infinite continuous mixture model, Dirichlet distribution, Bayesian estimation, asymptotic behavior, *a posteriori*.

1. Introduction

Dirichlet distribution is generally perceived as a family of continuous multivariate probability distributions, considered as a generalization of the Beta distribution. Reviewing the literature, the Dirichlet distribution appears in many fields, including the Bayesian analysis; compositional data modelling; multivariate analysis; statistical genetics; non-parametric inference; reliability; stochastic processes and so on [13, 18]. In Bayesian statistics, the Dirichlet distribution is the first choice and the most natural law used to model proportional data. It is used to describe the prior distributions in the Bayesian inference

*Corresponding Author.

Email addresses: sel_boukab@yahoo.com (S. Boukabour), affmasmoudi@gmail.com (A. Masmoudi)

Received: 04.10.2020; Accepted: 04.06.2021

and in other similar problems. This is due to the fact that the Dirichlet distribution is a conjugate law for counts following a multinomial distribution [12, 19].

The mixture of distributions has been in use to model data that have a multi-modal distribution with complex structure for many years in several areas of knowledge [17, 21]. The finite mixture of Dirichlet distributions was shown to be the most efficient and flexible law for modelling proportional data [2, 10, 15]. In their article [17], the authors emphasized the flexibility of finite mixtures considering them as a semiparametric model used to represent the shapes of unknown distributions. A crucial point for modelling this mixture is to determine the number of mixture components that best describes the database [9]. Indeed, it is not necessary to limit the components in the mixture models and assume that it is finite. Actually, the aim was to choose the appropriate distribution that best describes the data. Gaussian mixture models have been broadly used in various real applications. However, they can not always be considered as the best choice that well fits real data. In many settings, other distributions are applied, among which we can mention: the finite Dirichlet mixture model [2, 3, 5], finite generalized Dirichlet mixture model [4] and infinite mixture of Dirichlet distributions [9].

Nowadays, mixture models are related to such different fields as medicine, genetics, biometrics, and marketing, among others. Several different features of mixture distributions make them valuable in data modeling. Actually, statistical models which are based on mixture distributions capture numerous particular properties in real data analysis as multimodality, skewness, pattern recognition and kurtosis. For further information, we refer the reader to [7, 17]. Different existing Bayesian methods are used to learn parameters of mixture models. For instance in [6], the prior distributions of the Bayesian method are used to estimate parameters of the Gaussian mixture by maximizing the likelihood given the mixing coefficients. In addition, Gerlach et al. [11] provides a Bayesian analysis of dynamic mixture models using Markov chain Monte Carlo methods.

In the classical inference, the maximum likelihood (ML) estimator has been a popular option for proportional data. Assuming a multinomial sampling model: $(N_1, \dots, N_r) \sim Mult(N, \theta_1, \dots, \theta_r)$ where $N = \sum_{k=1}^r N_k$. It means that the multinomial likelihood function is given by $P(N_1 = n_1, \dots, N_r = n_r) = N! \prod_{k=1}^r \frac{\theta_k^{n_k}}{n_k!}$. The maximum likelihood estimator of the proportion θ_k is defined as $\hat{\theta}_k^{ML} = \frac{N_k}{N}$. The ML estimator has many optimal properties in statistics. It was shown that, under some regularity conditions, the ML estimator of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ is strongly consistent, unbiased and with a minimal variance. On the other hand, the estimation of proportions will be less efficient when the data is not big enough or the number of some cases in the data are equal to zero. Then, the need to add a prior knowledge is necessary. Assume that the parameters vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$, $0 < \theta_k < 1$ for all $k \in \{1, \dots, r\}$ and $\sum_{k=1}^r \theta_k = 1$, has a Dirichlet distribution with hyperparameters $(\alpha_1, \dots, \alpha_r)$ i.e. $\boldsymbol{\theta} \sim Dir(\alpha_1, \dots, \alpha_r)$. Since the Dirichlet distributions are conjugate prior for multinomial distributions, the *posterior* distribution of $(\theta_1, \dots, \theta_r) | N_1, \dots, N_r$ is also distributed as the Dirichlet with parameters $(N_1 + \alpha_1, \dots, N_r + \alpha_r)$ i.e. $(\theta_1, \dots, \theta_r) | N_1, \dots, N_r \sim Dir(N_1 + \alpha_1, \dots, N_r + \alpha_r)$. Thus, the expectation *a posteriori* (EAP) estimator of θ_k is given by

$$\hat{\theta}_k^{EAP} = \mathbb{E}(\theta_k | N_1, \dots, N_r) = \frac{N_k + \alpha_k}{N + \alpha}$$

where $\alpha = \sum_{k=1}^r \alpha_k$. In this work, the problem to be treated is: Suppose that two experts gave two different prior parameters α_k and α'_k for θ_k , even if we had the same proportions $\frac{\alpha_k}{\alpha} = \frac{\alpha'_k}{\alpha'}$, both estimators would not be equal i.e. $\frac{N_k + \alpha_k}{N + \alpha} \neq \frac{N_k + \alpha'_k}{N + \alpha'}$. The major contribution of this work was to assume that the sample size of the prior information α is a random variable, and to propose a new extension of the infinite mixture of Dirichlet distributions. As a natural mixing for continuous cases, we suggest taking an integral with respect to

the probability distribution of the random variable α instead of a sum. The family of distributions, we introduce, is considered without specifying a particular sample size α of the prior knowledge. It is called the infinite continuous mixture of Dirichlet distributions.

Therefore, the mixture expectation *a posteriori* (MEAP) estimator of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ was established using the infinite continuous mixture of Dirichlet laws as an alternative prior to the Dirichlet distribution. The resulting estimator was shown to have suitable properties such as the bias, variance, asymptotic normality and consistency. The performance of the proposed estimator is briefly studied through a comparative study between the MEAP and EAP estimators. This comparison was achieved by calculating the bias and variance of both estimators, in order to find out in which cases the proposed estimator has performed better than the other. The findings show that the behaviors of bias and variance are opposite to each other. According to a numerical simulation, we were able to conclude that the prior infinite mixture of Dirichlet distributions is the most flexible method for modelling data.

The rest of the paper is organized as follows. Section 2 provided a comprehensive background of the Dirichlet distribution, including its basic derivatives, finite and infinite mixtures of Dirichlet distributions. In Section 3, the infinite continuous mixture of Dirichlet distributions was established as an extension of the usual mixture of Dirichlet distributions. The mixture expectation *a posteriori* estimator was proposed in order to form an unrelated estimator of the prior sample size. Some asymptotic properties of the proposed estimator like bias, variance, consistency and asymptotic normality were established. Section 4 was devoted to the experimental results of the parameters estimation. The performance of the proposed MEAP estimator and the other classical ones (EAP, ML and Jeffreys estimators) were compared through a simulation study to illustrate our findings. Lastly, some conclusions and prospects were revealed in Section 5.

2. Background

2.1. Mixture of Dirichlet distributions

The Dirichlet distribution, which is constructed over multinomial space, is one of the most widely used multivariate distributions. This distribution is used in a variety of fields because of its ample flexibility and multidimensionality. In general, Dirichlet distributions are used as a prior knowledge. Knowing that the Dirichlet distribution is a conjugate probability distribution, the *posterior* distribution is also a Dirichlet distribution. A random vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ is said to have a Dirichlet distribution i.e. $\boldsymbol{\theta} \sim \mathcal{Dir}(\alpha_1, \alpha_2, \dots, \alpha_r)$ if

$$P(\boldsymbol{\theta}) = \Gamma\left(\sum_{k=1}^r \alpha_k\right) \prod_{k=1}^r \frac{\theta_k^{\alpha_k - 1}}{\Gamma(\alpha_k)}, \alpha_k > 0 \quad (2.1)$$

where $0 < \theta_k < 1$, for $k \in \{1, \dots, r\}$ and $\sum_{k=1}^r \theta_k = 1$, $(\alpha_1, \dots, \alpha_r)$ is a vector of the positive hyperparameters and $\alpha = \sum_{k=1}^r \alpha_k$. When all α_k ($k \in \{1, \dots, r\}$) are zeros, the Dirichlet distribution becomes noninformative [18]. It allows numerous symmetric and asymmetric modes [2]. Rather, it can be skewed to the right, skewed to the left or symmetric [5].

A finite mixture of Dirichlet distributions with M components is a convex combination of Dirichlet distributions; it is defined as

$$P(\boldsymbol{\theta}|\pi, \alpha^1, \dots, \alpha^M) = \sum_{j=1}^M \pi_j \mathcal{Dir}(\boldsymbol{\theta}|\alpha^j), \alpha^j = (\alpha_j^1, \dots, \alpha_j^r),$$

where $\alpha_j^k > 0$, for all $k \in \{1, \dots, r\}$ and $\pi_j > 0$ are the mixing coefficients or proportions with $\sum_{j=1}^M \pi_j = 1$ and $\mathcal{Dir}(\boldsymbol{\theta}|\alpha^j)$ is the Dirichlet distribution for the j^{th} component [2]. An important challenge in the finite mixture modelling is to decide the number of consistent mixtures that best describes the data [3]. Recent works have solved this problem using

several methods and algorithms in order to estimate the parameters and the number of mixing components. For example, Bouguil and Ziou [2] used an algorithm based on the Fisher scoring method for solving the issue of learning a finite Dirichlet mixture model. The infinite Dirichlet mixture model is also one of the best solutions to solve this problem by tending the number of mixing components M to infinity, see [9]. The model was achieved using a prior Dirichlet process for the model parameters. The stick-breaking construction is one of the most explicit and general Dirichlet processes. It is simply formulated as follows: Let $G \sim DP(s, H)$ represent a Dirichlet process where s is a concentration parameter and H is a base distribution. By adding the conditions of

$$G = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}, \quad \pi_j = \lambda_j \prod_{m=1}^{j-1} (1 - \lambda_m), \quad \lambda_j \sim \text{Beta}(1, s) \quad \text{and} \quad \theta_j \sim H \quad (2.2)$$

where δ_{θ_j} represents the Dirac measure. The mixing coefficients π_j are gotten by recursively breaking a unit length stick into an infinite number of pieces [9, 16, 20]. θ is assumed to be generated from a Dirichlet mixture. As it is also considered as a Dirichlet process using the stick-breaking construction, then the infinite Dirichlet mixture model (discrete case) is defined as

$$P(\theta | \pi, \alpha^1, \dots, \alpha^M) = \sum_{j=1}^{\infty} \pi_j \text{Dir}(\theta | \alpha^j). \quad (2.3)$$

Here, the assumption of $\sum_{j=1}^{\infty} \pi_j = 1$ should be achieved. The infinite mixture of Dirichlet distributions was shown as an extension of the finite mixture of Dirichlet ones. In what follows, we generalized the infinite Dirichlet mixture model, represented in Equation (2.3), from a discrete to a continuous case by proposing the new mixing that builds on an integral mixture.

3. Infinite continuous mixture of Dirichlet distributions

The problem to be treated here is how to choose the sample size α of the prior dataset in the Bayesian estimation. In what follows, we suggest changing variables $\alpha_k = \tau_k \alpha$. Noting that τ_k denotes a prior probability where $0 < \tau_k < 1$ and $\sum_{k=1}^r \tau_k = 1$. Thus, $\alpha = \sum_{k=1}^r \alpha_k$ is the prior sample size, which is supposed to be a random variable. The parameter τ_k represents the a priori proportion of observing the state k which has been evaluated from expert knowledge.

Let us define a random vector $\theta = (\theta_1, \dots, \theta_r)$ such that $0 < \theta_k < 1, \forall k = 1, \dots, r$ and $\sum_{k=1}^r \theta_k = 1$. Assume that the conditional probability distribution of θ given α is distributed as the following Dirichlet distribution:

$$\theta | \alpha \sim \text{Dir}(\tau_1 \alpha, \dots, \tau_r \alpha),$$

where $0 < \tau_k < 1$ and $\sum_{k=1}^r \tau_k = 1$. Suppose that α is a positive random variable with density function $\tilde{\pi}(\cdot)$, rather, $\tilde{\pi}$ is called the hyper-prior distribution. Then,

$$P(\theta | \alpha) = \Gamma(\alpha) \prod_{k=1}^r \frac{\theta_k^{\tau_k \alpha - 1}}{\Gamma(\tau_k \alpha)}.$$

It is easy to see that, $\mathbb{E}(\theta | \alpha) = \tau = (\tau_1, \dots, \tau_r)$, $\text{Var}(\theta_k | \alpha) = \frac{\tau_k(1-\tau_k)}{\alpha+1}$, for all $k = 1, \dots, r$ and $\text{Cov}(\theta_k, \theta_m | \alpha) = \frac{-\tau_k \tau_m}{\alpha+1}$. Statistically, if $\theta | \alpha \sim \text{Dir}(\tau_1 \alpha, \dots, \tau_r \alpha)$ and $\alpha \sim \tilde{\pi}(\cdot)$, then, θ follows the infinite continuous mixture of Dirichlet distributions with parameters vector

$\boldsymbol{\tau}$, i.e. $\boldsymbol{\theta} \sim \mathcal{MDir}(\tau_1, \dots, \tau_r)$; its density function is given by

$$\begin{aligned} P(\boldsymbol{\theta}) &= \int_0^\infty P(\boldsymbol{\theta}|\alpha)\tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \Gamma(\alpha) \left(\prod_{k=1}^r \frac{\theta_k^{\tau_k\alpha-1}}{\Gamma(\tau_k\alpha)} \right) \tilde{\pi}(\alpha) d\alpha. \end{aligned} \tag{3.1}$$

As the integral in expression (3.1) has a complex form, the numerical integration should be used. In order to approximate this integral, we suggest using Monte-Carlo method. This method is a powerful technique to approximate solutions to computational mathematical problems. It deals with approximating integrals which are made up of combinations between a function and a probability density, see [8]. Hence, the integral appears as a mathematical expectation of the Dirichlet function with respect to the random variable having the probability density $\tilde{\pi}$.

Relying on the Fubini-Tonelli theorem, the expectation of the random variable $\boldsymbol{\theta}$ is

$$\begin{aligned} \mathbb{E}(\boldsymbol{\theta}) &= \int_{\{\sum_{k=1}^r \theta_k=1\}} \boldsymbol{\theta} P(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_0^\infty \int_{\{\sum_{k=1}^r \theta_k=1\}} \boldsymbol{\theta} \left(\Gamma(\alpha) \prod_{k=1}^r \frac{\theta_k^{\tau_k\alpha-1}}{\Gamma(\tau_k\alpha)} \right) d\boldsymbol{\theta} \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty E(\boldsymbol{\theta}|\alpha) \tilde{\pi}(\alpha) d\alpha \\ &= (\tau_1, \tau_2, \dots, \tau_r) = \boldsymbol{\tau}. \end{aligned}$$

In the same way, the variance of the parameter θ_k is of the form

$$\text{Var}(\theta_k) = \int_0^\infty \frac{\tau_k(1 - \tau_k)}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha, \quad k = 1, \dots, r. \tag{3.2}$$

Moreover, the covariance of (θ_k, θ_m) is given by

$$\text{Cov}(\theta_k, \theta_m) = \int_0^\infty \frac{-\tau_k\tau_m}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha, \quad k \neq m. \tag{3.3}$$

For the proof of formulas (3.2) and (3.3), see Lemma 5.1 in the Appendix.

From a theoretical point of view, these last probabilistic properties (the expectation, variance and covariance of the random vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$) are independent from the choice of the prior sample size α , since the integration is achieved with respect to the parameter α . The following theorem provides the proposed estimator of the random vector $\boldsymbol{\theta}$ which follows the mixture of Dirichlet distributions $\mathcal{MDir}(\tau_1, \dots, \tau_r)$ and its expressions of bias and variance. We denote by $\tilde{\theta}_k^{MEAP}$ the mixture expectation *a posteriori* estimator of $\boldsymbol{\theta}$.

Theorem 3.1. *Let $V = (N_1, \dots, N_r)$ be a random vector following the multinomial distribution, i.e. $V \sim \mathcal{M}(N, \boldsymbol{\theta})$ where $N = \sum_{k=1}^r N_k$ and $\boldsymbol{\theta} \sim \mathcal{MDir}(\tau_1, \dots, \tau_r)$, then*

- (1) $\tilde{\theta}_k^{MEAP} = \int_0^\infty \hat{\theta}_k^{EAP}(\alpha) \tilde{\pi}(\alpha) d\alpha$, with $\hat{\theta}_k^{EAP}(\alpha) = \frac{N_k + \tau_k\alpha}{N + \alpha}$.
- (2) $\text{Bias}(\tilde{\theta}_k^{MEAP}) = (\tau_k - \theta_k) \left\{ 1 - \int_0^\infty \frac{N}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \right\}$.
- (3) $\text{Var}(\tilde{\theta}_k^{MEAP}) = N\theta_k(1 - \theta_k) \left\{ \int_0^\infty \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \right\}^2$.

The proof of Theorem 3.1 is given in the Appendix. In the case of the traditional approach of the expectation *a posteriori* estimator with the change of variables $\alpha_k = \tau_k\alpha$, we denote by α_0 a fixed sample size of the prior, then

$$\hat{\theta}_k^{EAP}(\alpha_0) = \frac{N_k + \tau_k\alpha_0}{N + \alpha_0} \quad \text{and} \quad \text{Bias}(\hat{\theta}_k^{EAP}(\alpha_0)) = (\tau_k - \theta_k) \left(1 - \frac{N}{N + \alpha_0} \right).$$

3.1. Performance evaluation of the MEAP estimator

In this subsection, we focused on the performance evaluation of the proposed MEAP and EAP estimators according to their biases and variances. The bias of an estimator is the difference between the estimator mean and the true value. Recall that, in the sense of bias, the proposed $\hat{\theta}_k^{MEAP}$ estimator is better than $\hat{\theta}_k^{EAP}$ when

$$\begin{aligned} \frac{|\text{Bias}(\hat{\theta}_k^{EAP}(\alpha_0))|}{|\text{Bias}(\hat{\theta}_k^{MEAP})|} &= \frac{1 - \frac{N}{N+\alpha_0}}{1 - \int_0^\infty \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} \geq 1 \Leftrightarrow \frac{\int_0^{+\infty} \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha - \frac{N}{N+\alpha_0}}{1 - \int_0^\infty \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} \geq 0 \\ &\Leftrightarrow \varphi(\alpha_0) = \int_0^{+\infty} \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha - \frac{N}{N+\alpha_0} \geq 0. \end{aligned} \quad (3.4)$$

Since $\alpha_0 \mapsto \varphi(\alpha_0)$ is a strictly increasing function with $\lim_{\alpha_0 \rightarrow 0} \varphi(\alpha_0) < 0$ and $\lim_{\alpha_0 \rightarrow +\infty} \varphi(\alpha_0) > 0$, thus

$$\varphi(\alpha_0) = 0 \Leftrightarrow \hat{\alpha}_0(N) = \frac{1}{\int_0^{+\infty} \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} - N. \quad (3.5)$$

Once $\alpha_0 \geq \hat{\alpha}_0(N)$, Equation (3.4) is satisfied. Then, we can say that the estimator $\tilde{\theta}_k^{MEAP}$ is better than $\hat{\theta}_k^{EAP}$ in terms of minimal bias.

In this proposed model, the a priori sample size α is supposed to be a random variable. Observe carefully that we only need to have a suitable support for α , then we assume that $\tilde{\pi}(\alpha)$ is a non-specific density prior and its support is $[\rho, +\infty[$, where $\rho > 0$ represents the lower bound of α . The following theorem establishes the asymptotic properties of the proposed MEAP estimator of θ_k for the infinite continuous mixture of Dirichlet distributions.

Theorem 3.2.

- (1) $\tilde{\theta}_k^{MEAP} \xrightarrow[N \rightarrow \infty]{a.s.} \theta_k$, almost surely.
- (2) $\sqrt{N} (\tilde{\theta}_k^{MEAP} - \theta_k) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, \theta_k(1 - \theta_k))$, in distribution.
- (3) The credible interval of θ_k is expressed by $\theta_k \in \left[\tilde{\theta}_k^{MEAP} \pm \frac{z_{1-\lambda/2} \hat{\sigma}_k}{\sqrt{N}} \right]$, where $z_{1-\lambda/2}$ is the percentile with the order $1 - \lambda/2$, $\lambda \in]0, 1[$ and $\hat{\sigma}_k = \sqrt{\tilde{\theta}_k^{MEAP}(1 - \tilde{\theta}_k^{MEAP})}$.
- (4) $\hat{\alpha}_0(N) \geq \rho$.
- (5) $\lim_{N \rightarrow +\infty} \hat{\alpha}_0(N) = \mathbb{E}(\alpha)$.

Proof.

- (1) Since $|\frac{N_k + \tau_k \alpha}{N + \alpha}| \leq 1$, then using the dominated convergence theorem, we have almost surely

$$\lim_{N \rightarrow \infty} \tilde{\theta}_k^{MEAP} = \int_0^\infty \lim_{N \rightarrow \infty} \frac{N_k + \tau_k \alpha}{N + \alpha} \tilde{\pi}(\alpha) d\alpha = \theta_k.$$

- (2) Performing the following assessment,

$$\begin{aligned} \sqrt{N} (\tilde{\theta}_k^{MEAP} - \theta_k) &= \int_0^\infty \sqrt{N} (\hat{\theta}_k^{EAP}(\alpha) - \theta_k) \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \sqrt{N} \left(\frac{N_k + \tau_k \alpha}{N + \alpha} - \theta_k \right) \tilde{\pi}(\alpha) d\alpha. \end{aligned}$$

If X_1, \dots, X_N is a sample taking values in $\{1, \dots, r\}$ with proportions $\theta_1, \dots, \theta_r$, respectively, then, $N_k = \sum_{i=1}^N \mathbb{1}_{\{X_i=k\}}$. Therefore,

$$\begin{aligned} \sqrt{N} (\tilde{\theta}_k^{MEAP} - \theta_k) &= \int_0^\infty \sqrt{N} \left(\frac{1}{N+\alpha} \sum_{i=1}^N \mathbb{1}_{\{X_i=k\}} + \frac{\tau_k \alpha}{N+\alpha} - \theta_k \right) \tilde{\pi}(\alpha) d\alpha \\ &= \sqrt{N} \left[\sum_{i=1}^N \mathbb{1}_{\{X_i=k\}} A_N - \theta_k \right] + \sqrt{N} \tau_k B_N \\ &= \sqrt{N} (N A_N f_N - \theta_k) + \sqrt{N} \tau_k B_N. \end{aligned} \quad (3.6)$$

Where $f_N = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_i=k\}} = \frac{N_k}{N}$, $A_N = \int_0^\infty \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha$ and $B_N = \int_0^\infty \frac{\alpha}{N+\alpha} \tilde{\pi}(\alpha) d\alpha$. Beginning with the first term of Equation (3.6),

$$\begin{aligned} \sqrt{N} (N A_N f_N - \theta_k) &= N A_N \sqrt{N} (f_N - \theta_k) + \theta_k \sqrt{N} (N A_N - 1) \\ &= N A_N \sqrt{N} (f_N - \theta_k) - \theta_k \sqrt{N} B_N. \end{aligned}$$

And using the central limit theorem, it is well known that $\sqrt{N} (f_N - \theta_k) \xrightarrow[N \rightarrow \infty]{D} N(0, \theta_k(1 - \theta_k))$. Also, $N A_N \xrightarrow[N \rightarrow \infty]{a.s.} 1$ and $\theta_k \sqrt{N} B_N \xrightarrow[N \rightarrow \infty]{a.s.} 0$. Then according to Slutsky's theorem, we have $\sqrt{N} (N A_N f_N - \theta_k) \xrightarrow[N \rightarrow \infty]{D} N(0, \theta_k(1 - \theta_k))$. The second term of (3.6): $\sqrt{N} \tau_k B_N \xrightarrow[N \rightarrow \infty]{a.s.} 0$. Applying Slutsky's theorem for the second time, we reach the desired result.

- (3) The credible interval of θ_k is deduced from the asymptotic normality of the proposed MEAP estimator.
- (4) Since the support of the prior density function $\tilde{\pi}$ is $[\rho, +\infty[$, then $\alpha \geq \rho$ and consequently

$$\frac{1}{\int_\rho^{+\infty} \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} - N \geq \frac{1}{\int_\rho^{+\infty} \frac{1}{N+\rho} \tilde{\pi}(\alpha) d\alpha} - N \Rightarrow \hat{\alpha}_0(N) \geq \rho.$$

- (5) According to Equation (3.5), $\lim_{N \rightarrow +\infty} \hat{\alpha}_0(N) = \mathbb{E}(\alpha)$. In fact,

$$\begin{aligned} \lim_{N \rightarrow \infty} \hat{\alpha}_0(N) &= \lim_{N \rightarrow +\infty} \left\{ \frac{N}{\int_\rho^{+\infty} \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} - N \right\} \\ &= \lim_{N \rightarrow +\infty} \left\{ \frac{N}{1 - \int_\rho^{+\infty} \frac{\alpha}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} - N \right\} \\ &= \lim_{N \rightarrow +\infty} \left\{ \frac{N}{1 - U_N} - N \right\}, \text{ where } U_N = \int_\rho^{+\infty} \frac{\alpha}{N+\alpha} \tilde{\pi}(\alpha) d\alpha \\ &= \lim_{N \rightarrow +\infty} \{N(1 + U_N + o(U_N)) - N\} \\ &= \int_\rho^{+\infty} \alpha \tilde{\pi}(\alpha) d\alpha \\ &= \mathbb{E}(\alpha). \end{aligned}$$

□
□

Herein, we focus on the assessment of comparing both variances of $\tilde{\theta}_k^{MEAP}$ and $\hat{\theta}_k^{EAP}$:

$$\begin{aligned} \text{Var}(\hat{\theta}_k^{EAP}(\alpha_0)) &= \text{Var}\left(\frac{N_k + \tau_k \alpha_0}{N + \alpha_0}\right) \\ &= \frac{1}{(N + \alpha_0)^2} \text{Var}(N_k) \\ &= \frac{N\theta_k(1 - \theta_k)}{(N + \alpha_0)^2}. \end{aligned}$$

See Theorem 3.1 for the $\text{Var}(\tilde{\theta}_k^{MEAP})$ formula. Then, we have the following equivalence

$$\begin{aligned} \frac{\text{Var}(\hat{\theta}_k^{EAP}(\alpha_0))}{\text{Var}(\tilde{\theta}_k^{MEAP})} &= \frac{\frac{1}{(N+\alpha_0)^2}}{\left(\int_0^\infty \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha\right)^2} - 1 \geq 0 \Leftrightarrow \frac{\frac{1}{(N+\alpha_0)^2} - \left(\int_0^\infty \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha\right)^2}{\left(\int_0^\infty \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha\right)^2} \geq 0 \\ &\Leftrightarrow \psi(\alpha_0) = \frac{1}{(N + \alpha_0)^2} - \left(\int_0^\infty \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha\right)^2 \geq 0. \end{aligned} \quad (3.7)$$

Since $\alpha_0 \mapsto \psi(\alpha_0)$ is a strictly decreasing function with $\lim_{\alpha_0 \rightarrow 0} \psi(\alpha_0) > 0$ and $\lim_{\alpha_0 \rightarrow +\infty} \psi(\alpha_0) < 0$, then

$$\psi(\alpha_0) = 0 \Leftrightarrow \hat{\alpha}_0(N) = \frac{1}{\int_0^{+\infty} \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha} - N. \quad (3.8)$$

Once $\alpha_0 \leq \hat{\alpha}_0(N)$, we can say that the estimator $\tilde{\theta}_k^{MEAP}$ outperforms the $\hat{\theta}_k^{EAP}$ in terms of minimal variance. Hence, the appropriate choice of the estimator has been shown to depend heavily on the $\hat{\alpha}_0(N)$ value, whether for the bias or the variance.

The expected squared deviation between the true value and the estimate, that is the mean squared error of both estimators $\hat{\theta}_k^{EAP}$ and $\tilde{\theta}_k^{MEAP}$ are represented, respectively, as follows:

$$\begin{aligned} \text{MSE}(\hat{\theta}_k^{EAP}) &= \text{Var}(\hat{\theta}_k^{EAP}) + \text{Bias}^2(\hat{\theta}_k^{EAP}) \\ &= \frac{N\theta_k(1 - \theta_k)}{(N + \alpha_0)^2} + (\tau_k - \theta_k)^2 \left(\frac{\alpha_0}{N + \alpha_0}\right)^2. \end{aligned} \quad (3.9)$$

$$\begin{aligned} \text{MSE}(\tilde{\theta}_k^{MEAP}) &= \text{Var}(\tilde{\theta}_k^{MEAP}) + \text{Bias}^2(\tilde{\theta}_k^{MEAP}) \\ &= N\theta_k(1 - \theta_k) \left(\int_0^{+\infty} \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha\right)^2 + (\tau_k - \theta_k)^2 \left(\int_0^{+\infty} \frac{\alpha}{N + \alpha} \tilde{\pi}(\alpha) d\alpha\right)^2. \end{aligned} \quad (3.10)$$

Of course, the MSE provides an effective measure of proximity. We are going to set a balance between variance and bias for Equations (3.9) and (3.10). We would minimize the bias in the MSE, remarking that when $N \rightarrow +\infty$ the bias is negligible compared to the variance as shown in both Equations (3.9) and (3.10). The remaining question is, therefore, whether we can keep a minimal variance if the estimator is asymptotically unbiased. In other words, even when the unbiased estimate sets the second term of these equations to 0, it may still have a very large MSE relative to other biased estimators. In this case, since the variance enlarges the credible interval, it should be as small as possible.

Before starting the simulation study, it is worth reminding the reader of Jeffreys noninformative prior [14]. The Jeffreys prior for θ is defined as $P(\theta) = |I(\theta)|^{1/2}$, where $|I(\theta)|$ is the determinant of the $r \times r$ Fisher information matrix. For multinomial distribution, the Jeffreys prior is given by $P(\theta) \propto \prod_{k=1}^r \theta_k^{-1/2}$. Then the Bayesian estimator of θ_k using the Jeffreys-rule prior is $\hat{\theta}_k^{Jef} = \frac{N_k + 1/2}{N + r/2}$. Interested readers may refer back for further details to [1] and [14].

4. Simulation study

In order to examine the performance of the proposed estimator $\tilde{\theta}_k^{MEAP}$, we began by generating a random variables vector $\theta = (\theta_1, \dots, \theta_6)$ which follows the mixture of Dirichlet distributions. That is to say, $\theta \sim \mathcal{MDir}(\tau)$, where $\tau = (0.08, 0.18, 0.15, 0.27, 0.19, 0.13)$. Let (N_1, \dots, N_r) be a random variable following the multinomial distribution $\mathcal{Mult}(N, \theta)$, where $N = 2000$. First, we took α which follows a truncated exponential probability density $\tilde{\pi}(\alpha) = \lambda e^{-\lambda(\alpha-\rho)} \mathbb{1}_{[\rho, +\infty[}(\alpha)$. In the second time, we considered the truncated Rayleigh density function $\tilde{\pi}(\alpha) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(\alpha-\rho)^2} \mathbb{1}_{[\rho, +\infty[}(\alpha)$.

We performed the maximum likelihood estimator $\hat{\theta}_k^{ML}$, the proposed one $\tilde{\theta}_k^{MEAP}$ with slightly inappropriate (unfavorable) priors τ and both Bayesian estimators of θ_k , $\hat{\theta}_k^{EAP}$ and $\hat{\theta}_k^{Jef}$. Using the Monte-Carlo method, we evaluated the proposed mixed Bayesian estimator $\tilde{\theta}_k^{MEAP}$. Let n be the number of samples, the performance of these estimators was assessed by means of the mean squared error ($MSE(\hat{\theta}) = \frac{1}{nr} \sum_{i=1}^n \sum_{k=1}^r (\hat{\theta}_k^i - \theta_k)^2$) between each of the three estimators and the true value of θ . Also, we set the mean absolute error ($MAE(\hat{\theta}) = \frac{1}{nr} \sum_{i=1}^n \sum_{k=1}^r |\hat{\theta}_k^i - \theta_k|$) of the compared estimators. The obtained numerical results are summarized in Figure 1 and Figure 2. These Figures show that the MSE and MAE for the mixture expectation a *posterior* estimator are clearly lower than those calculated with the classical EAP , Jef and ML estimators for both distributions of α (i.e. the truncated exponential and Rayleigh distributions).

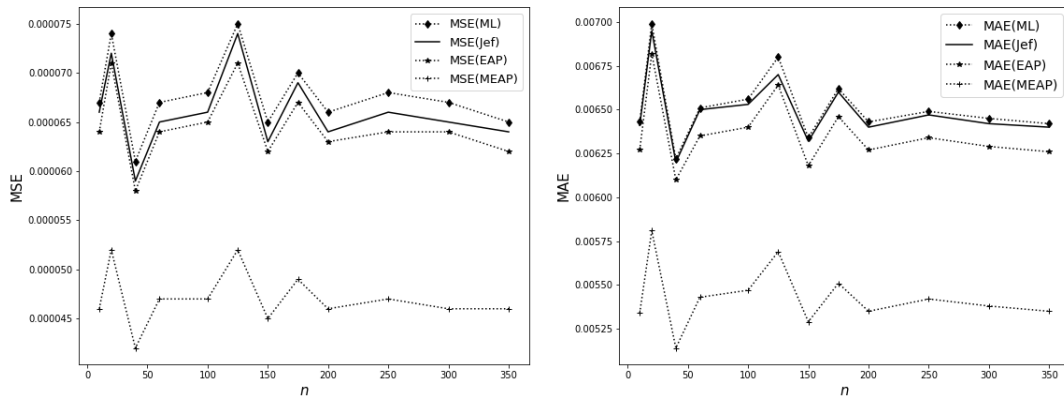


Figure 1. The variation of the MSE and MAE of the proposed MEAP, EAP and ML estimators for $\tilde{\pi}(\alpha) = 4e^{-4(\alpha-400)} \mathbb{1}_{[400, +\infty[}(\alpha)$ and $\alpha_0 = 50$.

According to Figures 3 and 4, we are able to highlight the behaviour of both functions φ and ψ (where $\varphi(\alpha_0) = \int_0^{+\infty} \frac{N}{N+\alpha} \tilde{\pi}(\alpha) d\alpha - \frac{N}{N+\alpha_0}$ and $\psi(\alpha_0) = \frac{1}{(N+\alpha_0)^2} - \left(\int_0^{+\infty} \frac{1}{N+\alpha} \tilde{\pi}(\alpha) d\alpha\right)^2$) which represent a comparison between the MEAP and EAP estimators of bias and variance, respectively. When these functions are positive, the bias absolute value and the variance for the proposed estimator are smaller than those obtained of the other one. These performances depend on the sample size N . The results shown in Figure 3 and Figure 4 correspond to the theoretical decision about comparing both absolute value of bias and variance of the $MEAP$ and EAP estimators. Actually, φ is an increasing function, Equation (3.4), and the proposed estimator is better in terms of bias when $\alpha_0 \geq \hat{\alpha}_0(N) \geq \rho$. Indeed, from $\alpha_0 \geq \rho = 400$, Figure 3, and $\rho = 100$ in Figure 4, the $|\text{Bias}(\hat{\theta}_k^{MEAP})|$ is clearly lower than $|\text{Bias}(\hat{\theta}_k^{EAP}(\alpha_0))|$. The same interpretation is used to describe the behaviour of ψ , which is a decreasing function and that the variance of the proposed estimator is better when $\alpha_0 \leq \rho$.

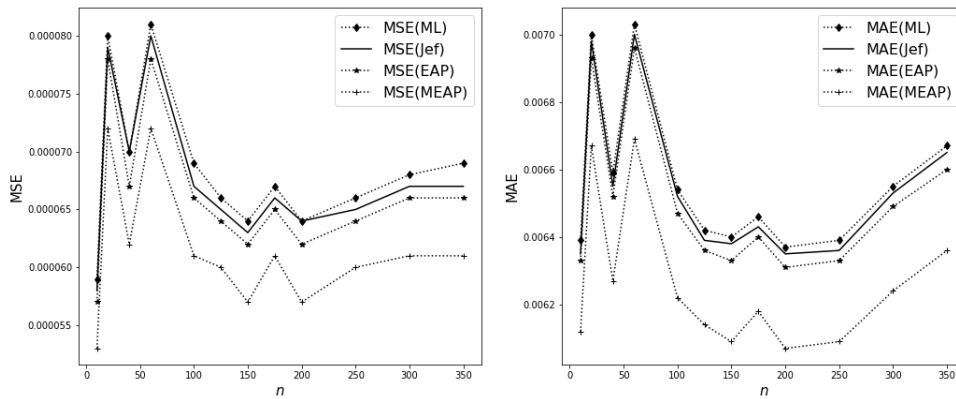


Figure 2. The variation of the MSE and MAE of the proposed MEAP, EAP and ML estimators for $\tilde{\pi}(\alpha) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(\alpha-100)^2} \mathbb{1}_{[100,+\infty[}(\alpha)$ and $\alpha_0 = 20$.

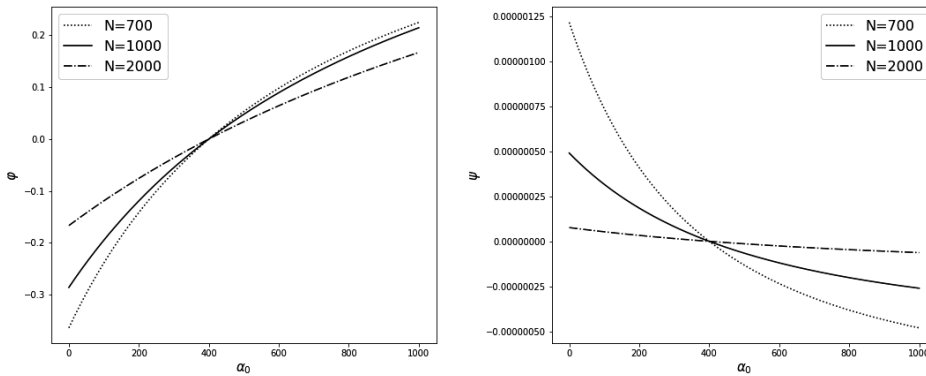


Figure 3. The variation of the functions φ and ψ in terms of α_0 with $\tilde{\pi}(\alpha) = 4e^{-4(\alpha-400)^2} \mathbb{1}_{[400,+\infty[}(\alpha)$ for several values of the prior sample size.

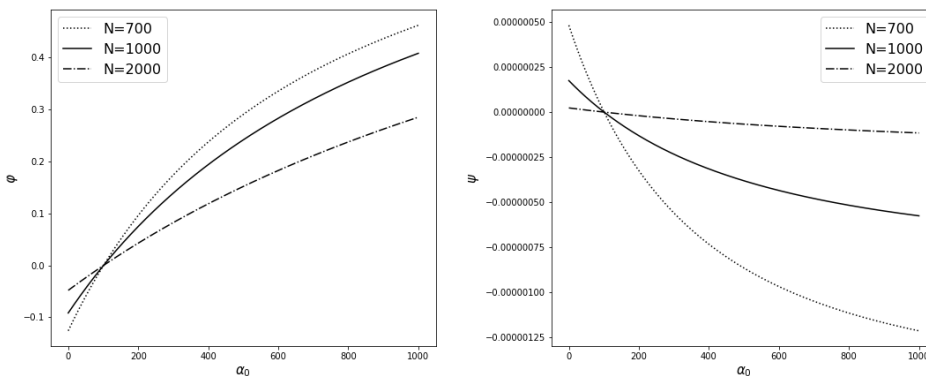


Figure 4. The variation of the functions φ and ψ in terms of α_0 with $\tilde{\pi}(\alpha) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(\alpha-100)^2} \mathbb{1}_{[100,+\infty[}(\alpha)$ for several values of the prior sample size.

Now, we fixed the number of repetitions n in order to visualize the behavior of the two estimators' MSE according to the variation of the fixed prior sample size α_0 . Obviously, the MSE of $\hat{\theta}_k^{MEAP}$ is independent from the value of α_0 . At this stage of analysis, we could assert that the proposed estimator performance is better than the classical EAP's, when α_0 is lower than the truncated $\rho = 300$, see Figure 5.

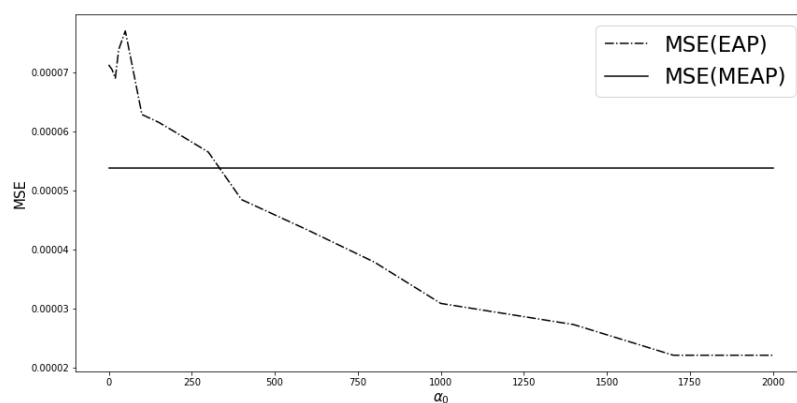


Figure 5. The variation of the MSE of both EAP and MEAP in terms of α_0 with $\tilde{\pi}(\alpha) = 0.7e^{-0.7(\alpha-300)} \mathbb{1}_{[300, +\infty[}(\alpha)$.

5. Conclusion

The widely used maximum likelihood estimator may suffer from a lack of performance when the number of discrete observations of some cases is zero and/or the size of data is small. Another interesting issue, the limitation of the usual Bayesian estimation is the choice of the prior sample size that changes the effectiveness of the expectation *a posteriori* estimator.

In this paper, we were interested in extending the mixture of Dirichlet distributions by proposing the infinite continuous mixture of Dirichlet distributions which represents a very crucial class of distributions in probability and statistical theory. We proposed a Bayesian estimator called the mixture expectation *a posteriori* estimator for this class of distributions. It allows us to have a unique estimator which is independent from the choice of parameters. Noteworthy, the asymptotic properties of this estimator such as bias, variance and asymptotic convergences were studied. Furthermore, bias and variance comparisons to the EAP estimator were achieved, confirming that the MEAP estimator outperforms the EAP one. We illustrated the estimation results by using two examples of simulation study. A comparative study between the proposed *MEAP*, the classical *EAP*, *Jef* and *ML* estimators based on the true values, was carried out. The results show a better precision and performance of the proposed *MEAP* estimator for the infinite continuous mixture model. In this work we dealt with a simulation study which achieved similar results as the theoretical findings. It offers different prospects for future research in an attempt to analyze and predict the data. The prospect of using real data has been evaluated taking into account this proposed mixture model as prior knowledge.

References

- [1] J.O. Berger, J.M. Bernardo and D. Sun, *Overall objective priors*, Bayesian Anal. **10** (1), 189-221, 2015.
- [2] N. Bouguila and D. Ziou, *Using unsupervised learning of a finite Dirichlet mixture model to improve pattern recognition applications*, Pattern Recognit. Lett. **26** (12), 1916-1925, 2005.
- [3] N. Bouguila and D. Ziou, *Online clustering via finite mixtures of Dirichlet and minimum message length*, Eng. Appl. Artif. Intell. **19** (4), 371-379, 2006.
- [4] N. Bouguila and D. Ziou, *High-dimensional unsupervised selection and estimation of a finite generalized Dirichlet mixture model based on minimum message length*, IEEE Trans. Pattern Anal. Mach. Intell. **29** (10), 1716-1731, 2007.

- [5] N. Bouguila, D. Ziou and J. Vaillancourt, *Novel mixtures based on the Dirichlet distribution: application to data and image classification*, in: International Workshop on Machine Learning and Data Mining in Pattern Recognition, 172-181, Springer.
- [6] C. Constantinopoulos, M.K. Titsias and A. Likas, *Bayesian feature and model selection for Gaussian mixture models*, IEEE Trans. Pattern Anal. Mach. Intell. **28** (6), 1013-1018, 2006.
- [7] V.H.L. Dávila, C.R.B. Cabral and C.B. Zeller, *Finite Mixture of Skewed Distributions*, Springer, 2018.
- [8] I.T. Dimov, *Monte Carlo Methods for Applied Scientists*, World Scientific, 2008.
- [9] W. Fan and N. Bouguila, *Infinite Dirichlet mixture models learning via expectation propagation*, Adv. Data Anal. Classif. **7** (4), 465-489, 2013.
- [10] W. Fan, N. Bouguila and D. Ziou, *Variational learning for finite Dirichlet mixture models and applications*, IEEE Trans. Neural Netw. Learn. Syst. **23** (5), 762-774, 2012.
- [11] R. Gerlach, C. Carter and R. Kohn, *Efficient Bayesian inference for dynamic mixture models*, J. Amer. Statist. Assoc. **95** (451), 819-828, 2000.
- [12] A. Ghribi and A. Masmoudi, *A compound Poisson model for learning discrete Bayesian networks*, Acta Math. Sci. **33** (6), 1767-1784, 2013.
- [13] D. Heckerman, *A tutorial on learning with Bayesian networks*, in: Learning in Graphical Models, 301-354, Springer, 1998.
- [14] H. Jeffreys, *An invariant form for the prior probability in estimation problems*, in: Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, **186** (1007), 453-461, The Royal Society London, 1946.
- [15] B. G. Lindsay, *Mixture models: theory, geometry and applications*, in: NSF-CBMS Regional Conference Series in Probability and Statistics, i-163, JSTOR, 1995.
- [16] N. Manouchehri, H. Nguyen, P. Koochemeshkian, N. Bouguila and W. Fan, *Online variational learning of Dirichlet process mixtures of scaled Dirichlet distributions*, Inf. Syst. Front. **22** (5), 1085-1093, 2020.
- [17] G.J. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, 2004.
- [18] K.W. Ng, G.L. Tian and M.L. Tang, *Dirichlet and Related Distributions: Theory, Methods and Applications*, John Wiley & Sons, 2011.
- [19] K. Sjölander, K. Karplus, M. Brown, R. Hughey, A. Krogh, I.S. Mian and D. Haussler, *Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology*, Bioinformatics **12** (4), 327-345, 1996.
- [20] Y.W. Teh, M.I. Jordan, M.J. Beal and D.M. Blei, *Hierarchical dirichlet processes*, J. Amer. Statist. Assoc. **101** (476), 1566-1581, 2006.
- [21] M. Zitouni, M. Zribi and A. Masmoudi, *Asymptotic properties of the estimator for a finite mixture of exponential dispersion models*, Filomat **32** (19), 6575-6598, 2018.

Appendix

This appendix contains certain proofs backing our theoretical results:

Lemma 5.1.

(1) The variance of θ_k is given by

$$\text{Var}(\theta_k) = \int_0^\infty \frac{\tau_k(1 - \tau_k)}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha, \quad k = 1, \dots, r.$$

(2) The covariance of θ_k, θ_m for each $k \neq m$ is

$$\text{Cov}(\theta_k, \theta_m) = \int_0^\infty \frac{-\tau_k \tau_m}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha, \quad k \neq m.$$

Proof. (1) Using the density function of θ and the Fubini-Tonelli theorem, then

$$\begin{aligned} \text{Var}(\theta_k) &= \int_{\{\sum_{k=1}^r \theta_k=1\}} \theta_k^2 P(\theta) d\theta - \mathbb{E}(\theta_k)^2 \\ &= \int_0^\infty \int_{\{\sum_{k=1}^r \theta_k=1\}} \theta_k^2 \left(\Gamma(\alpha) \prod_{k=1}^r \frac{\theta_k^{\tau_k \alpha - 1}}{\Gamma(\tau_k \alpha)} \right) d\theta \tilde{\pi}(\alpha) d\alpha - \tau_k^2 \\ &= \int_0^\infty \left(E(\theta_k^2 | \alpha) - E(\theta_k | \alpha)^2 \right) \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \text{Var}(\theta_k | \alpha) \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \frac{\tau_k(1 - \tau_k)}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha. \end{aligned}$$

(2) By the same way, the covariance of θ_k and θ_m where $k \neq m$ is

$$\begin{aligned} \text{Cov}(\theta_k, \theta_m) &= \int_{\{\sum_{k=1}^r \theta_k=1\}} \theta_k \theta_m P(\theta) d\theta - \mathbb{E}(\theta_k) \mathbb{E}(\theta_m) \\ &= \int_0^\infty \int_{\{\sum_{k=1}^r \theta_k=1\}} \theta_k \theta_m \left(\Gamma(\alpha) \prod_{k=1}^r \frac{\theta_k^{\tau_k \alpha - 1}}{\Gamma(\tau_k \alpha)} \right) d\theta \tilde{\pi}(\alpha) d\alpha - \tau_k \tau_m \\ &= \int_0^\infty \text{Cov}(\theta_k, \theta_m | \alpha) \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \frac{-\tau_k \tau_m}{\alpha + 1} \tilde{\pi}(\alpha) d\alpha. \end{aligned}$$

□

Proof of Theorem 3.1. (1) Evaluating the conditional expectation of θ_k given the observations $V = (N_1, \dots, N_r)$,

$$\begin{aligned} \tilde{\theta}_k^{MEAP} &= \int_{\{\sum_{k=1}^{r_i} \theta_k=1\}} \theta_k P(\theta | V) d\theta \\ &= \int_0^\infty \int_{\{\sum_{k=1}^{r_i} \theta_k=1\}} \theta_k \Gamma(N + \alpha) \left(\prod_{k=1}^{r_i} \frac{\theta_k^{N_k + \tau_k \alpha - 1}}{\Gamma(N_k + \tau_k \alpha)} \right) d\theta \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \frac{N_k + \tau_k \alpha}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \\ &= \int_0^\infty \hat{\theta}_k^{EAP}(\alpha) \tilde{\pi}(\alpha) d\alpha. \end{aligned}$$

(2) We begin with the bias of the classical estimator $\hat{\theta}_k^{EAP}(\alpha)$, with the change of variables $\alpha_k = \tau_k \alpha$

$$\begin{aligned} \text{Bias}(\hat{\theta}_k^{EAP}(\alpha)) &= \mathbb{E}(\hat{\theta}_k^{EAP}(\alpha)) - \theta_k \\ &= \frac{N_k + \tau_k \alpha}{N + \alpha} - \theta_k \\ &= \frac{N}{N + \alpha} \theta_k + \frac{\tau_k \alpha}{N + \alpha} - \theta_k \\ &= (\tau_k - \theta_k) \left(\frac{\alpha}{N + \alpha} \right). \end{aligned}$$

Using Fubini-Tonelli theorem, we get the following bias of $\tilde{\theta}_k^{MEAP}$.

$$\begin{aligned} \text{Bias}(\tilde{\theta}_k^{MEAP}) &= \mathbb{E}(\tilde{\theta}_k^{MEAP}) - \theta_k \\ &= \int_0^\infty \text{Bias}(\hat{\theta}_k^{EAP}(\alpha)) \tilde{\pi}(\alpha) d\alpha \\ &= (\tau_k - \theta_k) \int_0^\infty \left(\frac{\alpha}{N + \alpha} \right) \tilde{\pi}(\alpha) d\alpha \\ &= (\tau_k - \theta_k) \left\{ 1 - \int_0^\infty \frac{N}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \right\}. \end{aligned}$$

Observe that in the favourable case, indeed $\tau_k = \theta_k$, the $\tilde{\theta}_k^{MEAP}$ is an unbiased estimator.

(3) We now prove the variance formula of $\tilde{\theta}_k^{MEAP}$

$$\begin{aligned} \text{Var}(\tilde{\theta}_k^{MEAP}) &= \mathbb{E} \left((\tilde{\theta}_k^{MEAP})^2 \right) - \mathbb{E} \left(\tilde{\theta}_k^{MEAP} \right)^2 \\ &= \mathbb{E} \left(\left\{ N_k \int_0^\infty \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha + \tau_k \int_0^\infty \frac{\alpha}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \right\}^2 \right) - \mathbb{E} \left(\tilde{\theta}_k^{MEAP} \right)^2 \\ &= \mathbb{E} \left(\{N_k A_N + \tau_k B_N\}^2 \right) - \{\theta_k A_N + \tau_k B_N\}^2 \\ &= A_N^2 \mathbb{E}(N_k^2) + 2\tau_k A_N B_N \mathbb{E}(N_k) - (\theta_k N A_N)^2 - 2\theta_k \tau_k N A_N B_N \\ &= A_N^2 \left(N\theta_k(1 - \theta_k) + N^2\theta_k^2 \right) - (\theta_k N A_N)^2 \\ &= N\theta_k(1 - \theta_k) \left\{ \int_0^\infty \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha \right\}^2 \end{aligned}$$

where $A_N = \int_0^\infty \frac{1}{N + \alpha} \tilde{\pi}(\alpha) d\alpha$ and $B_N = \int_0^\infty \frac{\alpha}{N + \alpha} \tilde{\pi}(\alpha) d\alpha$. □