Smooth vs. likelihood estimation for a class of mixtures of discrete distributions

V. Patilea*   J.-M. Rolin†

December, 1995

Abstract

The problem of estimating the mixing distribution Q of a power series distributions (PSD's) mixture model is considered. In the first part of the paper some results on existence, support size, uniqueness and convergence of the (asymptotic) maximum likelihood estimator are presented. It is allowed for the mixture model to be misspecified.

In the second part of the paper we look at the nonparametric Bayesian approach for estimating mixtures. In this approach a Dirichlet random probability can be considered as a prior on the space of the mixing distributions. The form of the posterior distribution of the Dirichlet process it is known to be very complicated and by consequence the Bayesian estimators are not easily computable. However it has been remarked before that, conditioning on some additional information, the form of the nonparametric Bayesian estimator of the mixing distribution is much simpler. In a PSD's mixture model this remark gives an idea for a smooth estimator of the mixing distribution which has the form of a finite mixture of natural conjugate priors of the PSD's compound family. This smooth estimator is, in a sense, a kind of kernel estimator. Convergence results are obtained for this estimator. It turns out that, under quite mild conditions, the proposed estimator is an asymptotic maximum likelihood estimator. Some simulations show the behavior of this estimator for small and moderate sample sizes in the case of Poisson mixtures.

[1]Institute of Statistics, Louvain-la-Neuve

[2]Institute of Statistics and CORE, Louvain-la-Neuve

1 Introduction

About a hundred years ago K. Pearson introduced mixture models in order to study forehead breadth of Naples crabs. Since then mixtures of distributions have become appealing tools for more flexible and realistic statistical analysis. Among the mixture models, mixtures of discrete distributions have received extensive attention.

In order to introduce the mixture model we consider here let us denote by $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ a known parametric family of probability measures (p-measures) on the sample space $(\mathbf{X}, \mathcal{B})$. For all θ, let $p_\theta(\cdot)$ denote the density of F_θ w.r.t. a dominating σ-finite measure μ. The F_θ p-measures are referred to as compound distributions and the family $\mathcal{F}$ as the compound family. Let $(X_i, \theta_i)_{i=1}^n$ be random variables (r.v.'s) with values in $\mathbf{X} \times \Theta$ (defined on an abstract probability space $(\Omega, \mathcal{M}, P)$). The mixture model studied here can be stated as

$$\begin{gathered} \underset{i=1}{\overset{n}{\perp\!\!\!\perp}} (X_i, \theta_i) \\ X_i \mid \theta_i \rightsquigarrow F_{\theta_i} \\ \theta_i \rightsquigarrow Q \end{gathered} \tag{1.1}$$

where Q is a distribution on the parameter space Θ. The distribution Q is referred to as the mixing distribution. This distribution is unknown and has to be estimated without any further assumptions, that is nonparametrically. The data consist of the realizations of the X variables. From (1.1) we get that $X_1, ...X_n$ are independent r.v.'s distributed according to the mixed distribution denoted by F_Q. The F_Q distribution is characterized by $p_Q(\cdot) \overset{def}{=} \int_\Theta p_\theta(\cdot) Q(d\theta)$, its density w.r.t. μ.

In this paper the interest focus on the case where $\mathbf{X}$ is $I\!N_0 = \{0, 1, 2, ...\}$ and the compound family $\mathcal{F}$ consists of power series distributions (PSD's). The definition of a PSD's family will be recalled in section 2. In the same section, motivated by the paper of Simar (1976), we propose to estimate the mixing distribution Q using a nonparametric maximum likelihood type estimator. More precisely, we introduce (see also Pfanzagl (1988)) and we study some properties of the maximum likelihood estimator (MLE) and of the asymptotic maximum likelihood estimators (AMLE's) for PSD's mixtures. A AMLE $\hat{Q}_n$ is an estimator which almost maximizes the likelihood in the sense that the difference between the maximal value of the likelihood (as a function of Q) and the value of the likelihood corresponding to $\hat{Q}_n$ tends to zero as $n \to \infty$. The study of AMLE could be interesting, on one hand, for justifying numerical solutions for the ML estimator and, on the other hand, because one may prefer to consider estimators based on the likelihood principle with properties which are not shared by the MLE. We give results on existence, support size and uniqueness of the MLE. Afterwards the strong convergence of any sequence of AMLE's is proved. It is allowed for the mixture model to be misspecified, that is the true distribution generating the data may not be a power series distributions mixture; in such a case we assume the existence of a unique 'pseudo-true' mixing distribution. A quite weak sufficient condition for the existence of a such mixing distribution will be given.

In the third section we look at the nonparametric Bayesian approach for estimating mixtures as presented in Florens, Mouchart, Rolin (1992) (FMR (1992) from now on). In this framework a prior distribution, represented by a Dirichlet process, is introduced on the space of all p-measures on Θ. That is, for all i,

$$\begin{aligned} \theta_i \mid G &\rightsquigarrow G \\ G &\rightsquigarrow \mathcal{D}i(w_0 G_0) \end{aligned} \tag{1.2}$$

with G_0 a p-measure on $(\Theta, \mathcal{A})$ and $w_0 > 0$. FMR (1992) proved the consistency for the Bayesian estimators of functionals of the Dirichlet process G. However the form of the posterior distribution of the Dirichlet process G is very complicated and therefore the Bayesian estimators are not easily computable. FMR (1992) observed that, given the equality structure of the sample $\theta_1^n \stackrel{def}{=} (\theta_1, ..., \theta_n)$ of the Dirichlet process G, the form of the Bayesian estimators is much simpler. Unfortunately, given the observations $X_1^n \stackrel{def}{=} (X_1, ..., X_n)$, the equality structures have a practically unknown distribution and therefore it is not easy to generate them.

The idea proposed here is, in some sense, to look for the most likely equality structures of $\theta_1, ..., \theta_n$. This idea leads to an estimator of the mixing distribution which is a finite mixture of natural conjugate priors of the PSD's compound family. We will call this estimator a ENB (Empirical Nonparametric Bayesian) estimator of the mixing distribution. The properties of this estimator are studied in section 3. Contrary to the form of the MLE, the estimator of Q we propose is an absolutely continuous distribution. Therefore the ENB estimator offers a more realistic modeling for cases when the true mixing distribution is also an absolutely continuous distribution. We prove that the ENB estimator is a AMLE which, based on the results of section 2, assures the consistency of the ENB estimator.

Some empirical results and comparisons between the MLE and the ENB estimator for the case of Poisson mixtures are presented in the last part of the paper. In this case the ENB estimator has the form of a finite mixture of gamma distributions.

2 The (A)MLE in PSD's mixture models

As it was mentioned above the compound families $\mathcal{F}$ we consider in this paper consists of PSD's. In order to define a family of PSD's consider a function $a(\cdot) : [0, R] \mapsto (0, \infty]$, $a(\theta) \stackrel{def}{=} \sum_{k \geq 0} a_k \theta^k$, $a_k \geq 0$ for all $k \in I\!N_0$. It is supposed that R is the radius of convergence of the power series defining $a(\cdot)$, $0 < R \leq \infty$, and that $a_0 > 0$ (say $a_0 = 1$). Define for $\forall k \geq 0$ and $\forall \theta \in [0, R]$ the probabilities

$$p_\theta(k) = a_k\, \theta^k\, a(\theta)^{-1} \tag{2.1}$$

where $p_R(k) \stackrel{def}{=} \lim_{\theta\uparrow R} p_\theta(k) = 0$ if $a(R) = \infty$. Observe that the functions $\theta \mapsto p_\theta(k)$, $k \in I\!N_0$, are continuous on $[0, R]$. Denote $\Theta_R = [0, R]$ if $a(R)$ is finite and $\Theta_R = [0, R)$ otherwise. A family of PSD's defined by $a(\cdot)$ consists of distributions

$$F_\theta = \sum_{k\geq 0} p_\theta(k)\, \varepsilon_k \tag{2.2}$$

with $\theta \in \Theta_R$ and $p_\theta(k)$ as in (2.1) (the symbol ε_k stands for the p-measure concentrated at k). Remark that the mean of a r.v. distributed according to F_θ is $\theta a'(\theta)\, a(\theta)^{-1}$ and that for $\theta = 0$ we get in (2.2) $F_\theta = \varepsilon_0$. Some examples of PSD's are the Poisson distributions Po(θ), $a(\theta) = exp(\theta)$ and $R = \infty$, the negative binomial Nb$(v, 1-\theta)$, $v > 0$, $a(\theta) = (1-\theta)^{-v}$ and $R = 1$, the Hermite distribution of parameters (α, θ), $\alpha > 0$, $a(\theta) = exp(\alpha\,\theta + \theta^2/2)$ and $R = \infty$.

The compound families $\mathcal{F} = \{F_\theta,\ \theta \in \Theta\}$ we consider consists of PSD's F_θ as defined in (2.2). The parameter set Θ is a Borel subset of Θ_R. More precisely, for the sake of simplicity, for the rest of the paper Θ is an interval included in $[0, R]$: $\Theta = [c, \rho)$ if $a(\rho) = \infty$ and $\Theta = [c, \rho]$ otherwise. In what follows we consider $c = 0$ but the results can be adapted easily to the case $c > 0$.

In this section we study the MLE and the asymptotic MLE for mixtures of PSD's. First we assure that the parameter we want to estimate, i.e. the mixing distribution Q, is identifiable in the model we consider. For this, if not stated differently, we assume hereafter that the function $a(\cdot)$ defining the compound PSD's is such that :

H1: *There exists $b_1 > 0$, $b_l \geq 0$, $l \geq 2$, such that $b(\theta) \stackrel{def}{=} ln\, a(\theta) = \sum_{l\geq 1} b_l\, \theta^l$, $\forall\, \theta \in \Theta_R$.*

For a compound family of PSD's defined by $a(\cdot)$ as in *H1* it can be proved (see, for e.g., Lüxmann-Ellinghaus (1987)) that any element of $\mathcal{F}$ is an infinitely divisible distribution. The support of an infinitely divisible distribution on $I\!N_0$ is of the form $k_0 I\!N_0$, $k_0 \geq 1$. The condition $b_1 > 0$ is equivalent with $k_0 = 1$ and it does not represent a real restriction. It can be proved (see Lüxmann-Ellinghaus (1987)) that mixtures of distributions from a compound family $\mathcal{F}$ as above are identifiable, i.e. $\forall\, Q\ , Q'$, $F_Q = F_{Q'} \Rightarrow Q = Q'$. In the proof given by Lüxmann-Ellinghaus for the identifiability of PSD's mixtures the condition *H1* can be weakened (see also Milhaud, Mounime (1995)). However in this paper we mainly consider PSD's compound families $\mathcal{F}$ defined according to *H1*.

Some notations and definitions are introduced. For a sample $X_1, ..., X_n$ denote

$$M_n = \max_{1\leq i\leq n} X_i, \qquad p_{n,k} = n^{-1} \sum_{i=1}^{n} \mathbf{I}_{\{X_i = k\}}, \qquad K(n) = \sum_{k\geq 0} \mathbf{I}_{\{p_{n,k} > 0\}}$$

and $Y_1, ..., Y_{K(n)}$ the distinct data points. We say (see also Pfanzagl (1988)) that a sequence $(\hat{Q}_n)_{n\geq 1}$ of p-measures on Θ, defined for any $X_1^\infty \overset{def}{=} (X_1, X_2, ...)$, is a sequence of asymptotic maximum likelihood estimators (AMLE's) if there exists a measurable $B \subset \mathbf{X}^{I\!N_0}$ with $P(X_1^\infty \in B) = 1$, such that for all $X_1^\infty \in B$,

$$\lim_{n\to\infty} \sup_{Q\in\mathcal{P}(\Theta)} \sum_{k=0}^{\infty} p_{n,k}\, ln(p_Q(k)\,/p_{\hat{Q}_n}(k)) = 0 \tag{2.3}$$

(the $\mathcal{P}(\Theta)$ represents the space of all p-measures on Θ or, equivalently, the space of all p-measure Q on $I\!R$ with $Q(\Theta) = Q(I\!R) = 1$). If the supremum in (2.3) is zero for all n and for all possible X_1^n we obtain the usual definition of a sequence of MLE's.

The existence of a sequence of MLE's is proved in the following lemma (see also Van der Vaart, Wellner (1992)). The proof is given in the Appendix.

LEMMA 2.1 *Consider a PSD's compound family $\mathcal{F} = \{F_\theta,\, \theta \in \Theta\}$ not necessarily satisfying H1. For any sample X_1^n there exists a MLE $\hat{Q}_n$, i.e. there exists $\hat{Q}_n \in \mathcal{P}(\Theta)$ such that*

$$\sup_{Q\in\mathcal{P}(\Theta)} \sum_{k=0}^{\infty} p_{n,k}\, ln(p_Q(k)\,/p_{\hat{Q}_n}(k)) = 0. \tag{2.4}$$

A necessary tool for studying the properties of the MLE in mixture models is the directional derivative of the log-likelihood function $\bar{L}_n(Q) \overset{def}{=} n^{-1}\sum_{i=1}^n lnp_Q(X_i)$. Denote the directional derivative of $\bar{L}_n(\cdot)$ at Q toward Q' by $\bar{S}_n(Q, Q')$ and define it as

$$\bar{S}_n(Q, Q') = \lim_{\lambda\searrow 0} \lambda^{-1}\, [\, \bar{L}_n((1-\lambda)Q + \lambda Q') - \bar{L}_n(Q)\,]. \tag{2.5}$$

It can be proved that (see, for e.g., Lindsay (1983))

P1) $\bar{S}_n(Q, Q') = n^{-1}\sum_{i=1}^{n}(p_{Q'}(X_i)/p_Q(X_i)) - 1$

P2) $\hat{Q}_n$ is a MLE iff $\sup_{\theta\in\Theta} \bar{S}_n\,(\hat{Q}_n, \theta) = 0$ where $\bar{S}_n\,(Q, \theta) \overset{def}{=} \bar{S}_n\,(Q, \varepsilon_\theta)$

P3) given a MLE $\hat{Q}_n$, for all $\theta \in \Theta$ such that $\hat{Q}_n(\{\theta\}) > 0$, $\bar{S}_n(\hat{Q}_n, \theta) = 0$.

Assume that *H1* holds. Let $\hat{Q}_n$ be a MLE for which the existence was proved above. From he properties *P1* to *P3* stated above it turns out that, for any MLE $\hat{Q}_n$, $\bar{S}_n(\hat{Q}_n, \theta) < 0$, and consequently $\hat{Q}_n(\{\theta\}) = 0$, if $\theta \notin [\,\theta_m, \theta_M\,]$. The θ_m and θ_M are, respectively, the solutions of the equations $\theta\, b'(\theta) = (\zeta \wedge \min X_i)$ and $\theta\, b'(\theta) = (\zeta \wedge \max X_i)$, where $\zeta = \lim_{\theta\uparrow\rho} \theta b'(\theta)$ and ρ is the upper bound of the interval Θ ($x \wedge y$ denotes the minimum of the two numbers). In order to justify this statement remark that, for any $k \geq 0$, the function $\theta \to p_\theta(k)$ attains its maximum at θ_k, the solution of $\theta\, b'(\theta) = k$. If $\bar{S}_n(\hat{Q}_n, \theta) = 0$ for some $\theta \in \Theta$ which does not belong to

$[\,\theta_m, \theta_M\,]$, necessarily one of $\bar{S}_n(\hat{Q}_n, \theta_m)$ and $\bar{S}_n(\hat{Q}_n, \theta_M)$ is strictly positive, which is in contradiction with *P2*.

The next lemma will show that necessarily the equation $\bar{S}_n(\hat{Q}_n, \theta) = 0$ has a finite number of solutions θ. As $0 = \bar{S}_n(\hat{Q}_n, \hat{Q}_n) = \int_\Theta \bar{S}_n(\hat{Q}_n, \theta)\, \hat{Q}_n(d\theta)$, it turns out that $\hat{Q}_n$ should be concentrated at these solutions.

LEMMA 2.2 *Consider $b_l \geq 0$, $l \geq 1$ and $0 < \rho \leq \infty$ such that $\sum_{l\geq 1} b_l \theta^l$ converges for all $\theta \in [\,0, \rho)$. Take $t_1, ..., t_q > 0$ such that*

$$\sum_{h=1}^{q} t_h\, \theta^{i_h} \leq exp\left(\sum_{j\geq 1} b_j \theta^j\right) \tag{2.6}$$

for all $\theta \in [\,0, \rho)$ and some arbitrarly fixed $0 \leq i_1 < ... < i_q = N$.

Then there exists only a finite number r of distinct real values $\theta \in [\,0, \rho\,]$ satisfying (2.6) with equality and

$$\begin{array}{ll} r \leq (\,[\,\tilde{N}/2\,] + 1\,) \wedge q & \textit{if } \theta = 0 \textit{ satisfy (2.6) with equality} \\ r \leq [\,(\tilde{N}+1)/2\,] \ \wedge \tilde{q} & \textit{if not} \end{array}$$

where $\tilde{N} = N + 1$ if $\theta = \rho$ verifies (2.6) with equality and $a(R) < \infty$. Moreover, $\tilde{q} = q + 1$ if $\theta = \rho$ verifies (2.6) with equality and $i_1 > 0$. $\tilde{N} = N$, $\tilde{q} = q$ otherwise. ($[\,x\,]$ stands for the bigger integer smaller than x).

We will skip the proof of the previous lemma. It is based on the arguments that Simar (1976) used in his Lemma 3.1 for the case $[\,0, \rho) = [\,0, \infty)$ and $\sum_{l\geq 1} b_l \theta^l = \theta$. Apply the previous lemma for solving the equation $\bar{S}_n(\hat{Q}_n, \theta) = 0$ for a MLE $\hat{Q}_n$ in a mixture of PSD's from a compound family $\mathcal{F}$ as in *H1*. It turns out that any MLE is discrete, finitely supported, with the number of support points at most $(\,[\tilde{M}_n/2\,] + 1) \ \wedge \ K(n)$ if $\hat{Q}_n(\{0\}) > 0$ and at most $[\,(\tilde{M}_n + 1)\,/2\,] \ \wedge \ \tilde{K}(n)$ otherwise. $\tilde{M}_n = M_n + 1$ if ρ is finite, $\bar{S}_n(\hat{Q}_n, \rho) = 0$ and $a(R) < \infty$, $\tilde{M}_n = M_n$ otherwise. $\tilde{K}(n) = K(n) + 1$ if the observed sample does not contain zero, ρ is finite and $\bar{S}_n(\hat{Q}_n, \rho) = 0$, $\tilde{K}(n) = K(n)$ otherwise.

Given the distinct support points $\theta_1, ..., \theta_r$ that a MLE $\hat{Q}_n$ may have, the uniqueness of the MLE is equivalent with the uniqueness of $(\pi_1, ..., \pi_r)$ as solution of the equations

$$\sum_{s=1}^{r} \pi_s a_{Y_j} \theta_s^{Y_j}\, exp\left(-\sum_{l\geq 1} b_l \theta_s^l\right) = u_j^0, \ \ j = 1, ..., K(n), \tag{2.7}$$

where $(u_1^0, ..., u_{K(n)}^0)$ is given by (A.4) or by (A.6). This system has the $K(n) \times r$-matrix $(r \leq K(n) + 1)$ $A = (a_{j,s})$, $a_{j,s} = \theta_s^{Y_j}$ (read $\theta_s^{Y_j} = 1$ if $\theta_s = Y_j = 0$) and it has at least a solution (see lemma 2.1). If $r \leq K(n)$ the A matrix is of full rank column and the uniqueness of the MLE is obtained. If $r = K(n)+1$, which necessarily implies that one of the θ_s equals ρ, the MLE is not unique.

Finally, let say that $\hat{Q}_n = \sum_{j=1}^{r} \hat{\pi}_j \varepsilon_{\hat{\theta}_j}$. Remark that

$$\sum_{j=1}^{r} \hat{\pi}_j \hat{\theta}_j b'(\hat{\theta}_j) = n^{-1} \sum_{i=1}^{n} X_i \tag{2.8}$$

if all $\hat{\theta}_j$ belong to $[0, \rho)$ where ρ is the upper bound of the interval Θ (see, for e.g., Titterington, Smith, Makov(1985)).

We will talk about the convergence of the MLE in the more general framework of sequences of AMLE's. Considering AMLE's make sense, for instance, from the practical point of view. In general the MLE of the mixing distribution can only be approximated numerically. Moreover, the estimator proposed in section 3 turns out to be a AMLE. In what follows we use:

LEMMA 2.3 *Consider a compound family $\mathcal{F} = \{F_\theta,\, \theta \in \Theta\}$ of PSD's defined for a function $a(\cdot)$ not necessarily satifying H1. For any $\{Q_n\}_{n\geq 1} \subset \mathcal{P}(\Theta)$ and Q subprobability on $I\!R$,*

$$Q_n \to Q \ \textit{vaguely} \quad \Rightarrow \quad p_{Q_n}(k) \to p_Q(k),\ \forall k \in I\!N_0.$$

Proof: It suffices to justify the implication for $\Theta = \Theta_R$. We recall that, $\forall k \in I\!N_0$, $\theta \mapsto p_\theta(k)$ is continuous on Θ_R. If Θ_R is a compact interval necessarily $Q \in \mathcal{P}(\Theta)$. If Θ_R is not a compact interval $\lim_{\theta \uparrow R} p_\theta(k) = 0$. Thus $Q_n \to Q$ vaguely implies $p_{Q_n}(k) = \int_{\Theta_R} p_\theta(k)\, Q_n(d\theta) \to p_Q(k) = \int_{\Theta_R} p_\theta(k)\, Q(d\theta), \forall k$.

Q.E.D.

Define $\mathbf{q} = (q_k)_{k\geq 0}$, $q_k \stackrel{def}{=} P(X_1 = k)$, the true distribution generating the data. It is not necessarily assumed here that $\mathbf{q}$ represents a mixture of PSD's. Thus we will also study the limit of a sequence of AMLE's when the considered PSD's mixture model is misspecified. However we assume

H2: *There exists a unique $Q_0 \in \mathcal{P}(\Theta)$ such that*

$$\liminf_{m\to\infty} \sum_{k=0}^{m} q_k\, ln\,(p_Q(k)/p_{Q_0}(k)) < 0, \quad \forall Q \in \mathcal{P}(\Theta),\ Q \neq Q_0.$$

Observe that *H2* is automatically fullfilled if the mixture model is identifiable and well specified, i.e. if $\mathbf{q}$ represents a mixture of PSD's from the considered compound family. Denote $l(Q) = \sum_{k\geq 0} q_k\, ln\, p_Q(k)$ and $\mathcal{F}_{\mathbf{q}} = \{Q \in \mathcal{P}(\Theta) \mid l(Q) > -\infty\}$. Remark that $\mathcal{F}_{\mathbf{q}}$ is a convex set. The next lemma shows that $\mathcal{F}_{\mathbf{q}} \neq \emptyset$ implies *H2*, whether the model is well specified or not, provided that the mixtures are identifiable. It also gives a sufficient condition for all the mixtures models considered here and an equivalent condition for the case of Poisson mixtures for the nonemptiness of $\mathcal{F}_{\mathbf{q}}$. The proof of the next lemma is given in the Appendix.

LEMMA 2.4 *a) If the mixtures are identifiable $\mathcal{F}_{\mathbf{q}} \neq \emptyset$ implies the condition H2.*
b) For a compound family $\mathcal{F}$ satisfying H1

$$E_{\mathbf{q}}\left[X_1 ln(X_1+1)\right] < \infty \Rightarrow \mathcal{F}_{\mathbf{q}} \neq \emptyset \tag{2.9}$$

c) For Poisson mixtures models

$$E_{\mathbf{q}}\left[ln(X_1+1)\right] < \infty \Leftrightarrow \mathcal{F}_{\mathbf{q}} \neq \emptyset. \tag{2.10}$$

We showed that, even if the mixture model is wrong, a quite mild condition implies *H2*. However it seems that *H2* does not necessarily imply $\mathcal{F}_{\mathbf{q}} \neq \emptyset$. Finally the strong convergence of any AMLE's sequence is proved in the next proposition (see also Pfanzagl (1988)). This convergence will be obtained using only the condition *H2*. We study also the limit of the log-likelihood $\bar{L}_n(\hat{Q}_n)$ if $\{\hat{Q}_n\}_{n\geq 1}$ is an AMLE's sequence.

PROPOSITION 2.5 *a) Consider a mixture model of a compound family of PSD's defined for a function $a(\cdot)$ not necessarily satisfying H1. Assume that H2 holds. Then any sequence $\{\hat{Q}_n\}_{n\geq 1}$ of AMLE's is strongly convergent, i.e. with probability one $\hat{Q}_n \to Q_0$ weakly, where Q_0 is given by H2.*
b) If $\mathcal{F}_{\mathbf{q}} \neq \emptyset$ and $\{\hat{Q}_n\}_{n\geq 1}$ is an AMLE's sequence, with probability one,

$$\lim_{n\to\infty} \sum_{k\geq 0} p_{n,k}\, ln\left[\frac{p_{\hat{Q}_n}(k)}{p_{Q_0}(k)}\right] = 0. \tag{2.11}$$

where Q_0 is given by H2.

Proof: a) From the SLLN there exists $A \subset \mathbf{X}^{I\!N_0}$, $P(X_1^\infty \in A) = 1$, such that $p_{n,k} \to q_k$, $\forall k \geq 0$, for any $X_1^\infty \in A$. Let $\{\hat{Q}_n\}_{n\geq 1} \subset \mathcal{P}(\Theta)$ be an AMLE's sequence. For the rest of the proof of this proposition we consider, implicitly, only X_1^∞ from A for which (2.3) holds for $\{\hat{Q}_n\}_{n\geq 1}$. From the sequence $\{\hat{Q}_n\}_{n\geq 1}$ we can extract a subsequence $\{\hat{Q}_{n_j}\}_{n_j\geq 1}$ vaguely convergent to $\tilde{Q}$, a subprobability on $I\!R$. Lemma 2.3 tells that $p_{\hat{Q}_{n_j}}(k) \to p_{\tilde{Q}}(k) = \int_\Theta p_\theta(k)\, \tilde{Q}(\,d\theta)$, $\forall k \geq 0$. From (2.3) we get, for arbitrarly fixed $\alpha \in (0,1)$,

$$\limsup_{n_j\to\infty} \sum_{k\geq 0} p_{n_j,k}\, ln\left[1 + \alpha\left(\frac{p_{Q_0}(k)}{p_{\hat{Q}_{n_j}}(k)} - 1\right)\right] \leq 0. \tag{2.12}$$

For all k and n_j, $b_{n_j,k} \stackrel{def}{=} ln\,[1+\alpha((p_{Q_0}(k)/p_{\hat{Q}_{n_j}}(k)-1)\,]-ln(1-\alpha) \geq 0$. Thus $\forall\epsilon > 0$ there exists $N_\epsilon \in I\!N_0$ such that $\forall L \in I\!N_0$ and $\forall n_j \geq N_\epsilon$,

$$\sum_{k=0}^{L} p_{n_j,k} b_{n_j,k} \leq \sum_{k\geq 0} p_{n_j,k} b_{n_j,k} \leq -ln(1-\alpha)+\epsilon. \tag{2.13}$$

Letting first n_j and afterwards L to tend to infinity we obtain

$$\sum_{k\geq 0} q_k\, ln\left[1+\alpha\left(\frac{p_{Q_0}(k)}{p_{\tilde{Q}}(k)}-1\right)\right] \leq 0. \tag{2.14}$$

From the concavity of the logarithmic function we have :

$$0 \geq \sum_{k\geq 0} q_k\, ln\left[1+\alpha\left(\frac{p_{Q_0}(k)}{p_{\tilde{Q}}(k)}-1\right)\right] \geq \alpha \limsup_{m\to\infty} \sum_{k=0}^{m} q_k\, ln\,\frac{p_{Q_0}(k)}{p_{\tilde{Q}}(k)} \tag{2.15}$$

It turns out from *H2* that necessarily $\tilde{Q} = Q_0$. As, for almost all X_1^∞, any convergent subsequence of $\{\hat{Q}_n\}_{n\geq 1}$ has the weak limit Q_0, the strong convergence of any AMLE's sequence is proved.

b) First remark that as n increases to infinity,

$$\sum_{k\geq 0} p_{n,k}\, ln\, p_{Q_0}(k) \longrightarrow \sum_{k\geq 0} q_k\, ln\, p_{Q_0}(k) = l(Q_0) > -\infty. \tag{2.16}$$

It remains to prove that

$$\bar{L}_n(\hat{Q}_n) \longrightarrow l(Q_0). \tag{2.17}$$

for a (exact) MLE sequence $\{\hat{Q}_n\}_{n\geq 1}$ and afterwards to use (2.3). It easy to see that

$$\bar{L}_n(\hat{Q}_n) \leq \sum_{k=0}^{L} p_{n,k}\, ln\, p_{\hat{Q}_n}(k),\ \ \forall L \in I\!N \tag{2.18}$$

(because $ln\, p_{\hat{Q}_n}(k) \leq 0, \forall k \geq 0$) and that

$$\sum_{k\geq 0} p_{n,k}\, ln\, p_{Q_0}(k) \leq \sum_{k\geq 0} p_{n,k}\, ln\, p_{\hat{Q}_n}(k),\ \ \forall n \geq 1 \tag{2.19}$$

(because $\hat{Q}_n$ is a MLE). From (2.16) and (2.19) we get that $l(Q_0) \leq \liminf_{n\to\infty} \bar{L}_n(\hat{Q}_n)$. On the other hand from Lemma 2.3 and (2.18) we have that $\limsup_{n\to\infty} \bar{L}_n(\hat{Q}_n) \leq \sum_{k=0}^{L} q_k\, ln\, p_{Q_0}(k),\ \forall L \in I\!N_0$. Finally let L to increase to infinity and *b)* is proved.

Q.E.D.

3 Smooth estimation of PSD's mixtures

In a nonparametric Bayesian approach for mixture models a prior distribution represented by a Dirichlet process might be considered on $\mathcal{P}(\Theta)$. In this approach a mixture model can be written as

$$\begin{gathered}\coprod_{i=1}^{n}(X_i,\theta_i)\mid G \\ X_i\mid\theta_i \rightsquigarrow F_{\theta_i} \\ \theta_i\mid G\rightsquigarrow G \\ G\rightsquigarrow \mathcal{D}i(w_0G_0)\end{gathered} \tag{3.1}$$

with G_0 a p-measure on Θ and $w_0>0$. The p-measure G_0 can be interpreted as a prior guess and w_0 as a degree of confidence associated with this prior guess.

FMR (1992) studied the convergence properties of functionals of G in a (well specified) mixture model as in (3.1). Provided that the mixture is identifiable, they proved that, for any positive measurable function $f(\cdot)$ defined on Θ, $P-$a.s.,

$$\lim_{n\to\infty} E\left[\int_\Theta f(\theta)\,G(d\theta)\mid X_1^n\right]=\int_\Theta f(\theta)\,G(d\theta). \tag{3.2}$$

Unfortunately the Bayesian estimators are very difficult to compute. Following FMR (1992), assume that the equality structure of the (unobserved) sample θ_1^n is given. That is, considering $\mu_1,...,\mu_{L(n)}$ the distinct values of θ_1^n and $\Delta_i=l$ if $\theta_i=\mu_l$, it is assumed that $\Delta_1^n\overset{def}{=}(\Delta_1,...,\Delta_n)$ is given. In this case it can be observed that the form of the nonparametric Bayesian estimators becomes much simpler. We apply this remark for PSD's mixtures models defined acording to *H1*. On Θ we consider a family of natural conjugate priors $D(u,s)$, $u,s>0$, given by the densities of the form

$$m(\theta\mid u,s)=c_{u,s}\,\theta^{\,s-1}\,exp(-u\,b(\theta)),\qquad c_{u,s}^{-1}=\int_\Theta \theta^{\,s-1}\,exp(-u\,b(\theta))\,d\theta. \tag{3.3}$$

Remark that, for $\Theta=\Theta_R$ and $\lim_{\theta\uparrow R} b(\theta)=\infty$

$$c_{u,s}^{-1}=\frac{u}{s}\int_\Theta b'(\theta)\,\theta^{\,s}\,exp(-u\,b(\theta))\,d\theta. \tag{3.4}$$

Assume that G_0 represents the distribution $D(u_0,s_0)$, for some $u_0,s_0>0$, and denote $\nu_{n,l}=\sum_{i=1}^n \mathbf{I}_{\{\Delta_i=l\}}$ and $S_{n,l}=\sum_{\{i\mid\Delta_i=l\}}X_i$. Then, for every positive measurable function $f(\cdot)$

$$E\left[\int_\Theta f(\theta)\,G(d\theta)\mid \Delta_1^n,X_1^n\right]=\int_\Theta f(\theta)\,m(\theta\mid\Delta_1^n,X_1^n)\,d\theta \tag{3.5}$$

where $m(\theta \mid \Delta_1^n, X_1^n)$ is the density of the p-measure

$$\frac{w_0}{w_0+n}\, D(u_0, s_0) + \frac{n}{w_0+n} \sum_{l=1}^{L(n)} \frac{\nu_{n,l}}{n}\, D(u_0+\nu_{n,l}, s_0+S_{n,l}). \tag{3.6}$$

Even if (3.6) has a much simpler form, the problem is how to generate in practice, in a feasible way, the equality structures Δ_1^n in order to integrate out these variables. Some solutions using Markov chain Monte Carlo methods have been proposed (see MacEachern, Müller (1994) for a very recent reference).

In this paper we follow another idea inspired by (3.5) and (3.6). We look for the most likely equality structures Δ_1^n. In other words, we introduce an estimator for the mixing distribution of the form (3.6) corresponding, in fact, to the most likely $\nu_{n,l}$ and $S_{n,l}$. Moreover the w_0 can be taken zero, which corresponds, in a sense, to a noninformative prior on the space of the mixing distribution (nevertheless the convergence results proved below still hold for an estimator inspired from (3.6) corresponding to a positive w_0). Contrary to the form of the MLE, the obtained estimator is smooth.

Following these ideas define, for some $u_0, s_0 > 0$, a set of p-measures on Θ

$$\mathcal{P}_{L(n),t_n} \stackrel{def}{=} \{\, Q = Q_\psi = \sum_{l=1}^{L(n)} \pi_l\, D(u_0+n\pi_l, s_0+n\pi_l\theta_l b'(\theta_l)) \mid \sum_{l=1}^{L(n)} \pi_l = 1\} \tag{3.7}$$

where $\psi = (\pi_1, ..., \pi_{L(n)}, \theta_1, ..., \theta_{L(n)}) \in S \stackrel{def}{=} [\,0,1\,]^{L(n)} \times [\,0,t_n\,]^{L(n)}$, $[\,0,t_n\,] \subset \Theta$ and $L(n) \in \{1,2,...\}$. The estimator we propose here for mixtures of distributions from the family $\mathcal{F}$ is obtained by maximizing the log-likelihood function $\bar{L}_n(\cdot)$, restricted to $\mathcal{P}_{L(n),t_n}$. We call this estimator an ENB estimator. Thus

$$\hat{Q}_{ENB} \stackrel{def}{=} \arg \max_{Q \in \mathcal{P}_{L(n),t_n}} \bar{L}_n(Q) = \arg \max_{\psi \in S} \bar{L}_n(Q_\psi). \tag{3.8}$$

It is clear that, for all n, $\hat{Q}_{ENB}$ exists since S is a compact set and the log-likelihood function is continuous as a function of ψ. It can be remarked that $\mathcal{P}_{L(n),t_n}$ is not a convex set. Therefore it seems that the estimator defined in (3.8) is not unique.

As an example, in the case of Poisson mixtures with $\Theta = \Theta_R = [\,0,\infty)$, the ENB estimator we propose here has the form of a finite mixture of gamma distributions. For $Q = \sum_{l=1}^{L(n)} \pi_l\, \Gamma(u_0+n\pi_l,\, s_0+n\pi_l\theta_l)$ the F_Q distribution has the form of a finite mixture of negative binomial distributions, that is

$$Po_Q = \sum_{l=1}^{L(n)} \pi_l\, Nb\left(s_0 + n\,\pi_l\,\theta_l,\, \frac{u_0+n\,\pi_l}{1+u_0+n\,\pi_l}\right), \tag{3.9}$$

where $P(\,Y = k\,) = (\Gamma(v+k)/(\Gamma(v)\,k!))\,p^v(1-p)^k$, $\forall k \in I\!N_0$, if $Y \rightsquigarrow Nb(v,p)$, $p \in (\,0,1\,]$. The log-likelihood can be easily written for this case. However, in general, a numerical integration is needed for computing the $p_Q(k)$ quantities involved in the likelihood.

As a choice for $L(n)$ we propose the upper bound of the support size of the MLE given in Lemma 2.2. This choice agrees with the intuition that the estimator we propose here is, in a sense, a smoothed MLE with the $D(u,s)$ distribution playing the role of a kernel. The role of t_n is to assure the existence of the ENB estimator. When Θ is compact $t_n = \rho$, $\forall n$, could be a choice in the definition (3.7) (ρ denotes the upper bound of the interval Θ). In the case where the compound family $\mathcal{F}$ represents all the family $\mathcal{P}$ of PSD's defined by a function $a(\cdot)$ and $\Theta_R = [0, R)$, (3.6) may suggest a choice for t_n (and for the domain of π_l). As in the definition (3.7) $n\pi_l$ and $n\pi_l\theta_l b'(\theta_l)$ replace, respectively, $\nu_{n,l}$ and $\tilde{X}_{n,l}$ of (3.6), one may propose, for instance, to define $\mathcal{P}_{L(n),t_n}$ for $\pi_l \in \{0\} \cup [n^{-1}, 1]$ (or even for $\pi_l \in \{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$) and t_n such that $t_n b'(t_n) \leq \sum_{i=1}^{n} X_i$. Remark that the results below (mainly the lemma 3.1 *c)*) can easily be adapted for $\mathcal{P}_{L(n),t_n}$ defined with one of the previous restrictions on π_l.

In what follows we will prove the convergence of the ENB estimator defined in (3.8). Some preliminary results are needed. The point *c)* of the next lemma assures that any probability on Θ is a weak limit of finite mixtures of distributions as those defining $\mathcal{P}_{L(n),t_n}$. On the other hand the proof of *b)* could be, in fact, the proof of the consistency of the (parametric) Bayesian estimators for PSD's models defined acording to *H1*. The proof of the next lemma is given in the Appendix.

LEMMA 3.1 *a) Assume that in (3.3)* $s = s_n$, $u = u_n$, $u_n \to \infty$ *and* $s_n/u_n \to \gamma \in I\!R_+$. *Then*

$$[c_{u_n,s_n}]^{-\frac{1}{u_n}} \to \max_{\theta\in\Theta} \theta^{\gamma} exp(-b(\theta)). \tag{3.10}$$

b) Under the same hypothesis on s_n *and* u_n

$$D(d_n, c_n) \to \varepsilon_{\theta(\gamma)}, \ \ weakly, \tag{3.11}$$

where $\theta(\gamma) = \arg\max_{\theta\in\Theta} \theta^{\gamma} exp(-b(\theta))$.

c) Consider $\{L_n\}_{n\geq 1}$, $\{t_n\}_{n\geq 1}$ *such that* $L_n \to \infty$, $t_n \uparrow \rho$, $[0, t_n] \subset \Theta$. *Then*

$$\mathcal{G} \stackrel{def}{=} \bigcup_{n\geq 1} \mathcal{P}_{L(n),t_n} \tag{3.12}$$

is dense in the space of all probabilities on Θ *considered with the weak topology.*

Thus, for any $Q \in \mathcal{P}(\Theta)$, provided that $L(n) \to \infty$ and $t_n \uparrow \rho$, there exists $\{\check{Q}_n\}_{n\geq 1} \subset \mathcal{G}$, $\check{Q}_n \in \mathcal{P}_{L(n),t_n}$ for all n, such that $\check{Q}_n \to Q$ weakly. The following result studies the limit of the $\bar{L}_n(\check{Q}_n)$. The proof of the next lemma is also given in the Appendix.

LEMMA 3.2 *Assume that H1 holds, that* $E_{\mathbf{q}}[X_1 ln\,(X_1+1)] < \infty$ *and that* $\{\check{Q}_n\}_{n\geq 1} \subset \mathcal{G}$, $\check{Q}_n \to \check{Q}$ *weakly. If* $\check{Q} \neq \varepsilon_0$ *then,* $P-$a.s.,

$$-\infty < l(\check{Q}) = \lim_{n\to\infty} \sum_{k\geq 0} p_{n,k}\, ln\, p_{\check{Q}_n}(k). \tag{3.13}$$

Now we can prove the strong convergence of a sequence of ENB estimators . We obtain this convergence by proving that a ENB estimator is an AMLE. A 'classic' proof of the consistency for the MLE seems not applicable for the ENB estimator since $\bar{L}_n(\hat{Q}_{ENB}) \geq \bar{L}_n(Q_0)$ need not necessarily hold. For the next proof we denote by $\hat{Q}_{MLE}$ the MLE of a PSD's mixture.

PROPOSITION 3.3 *Consider a mixture model of a compound family of PSD's satifying H1. If*

$$E_{\mathbf{q}}[\,X_1 ln(X_1+1)\,] < \infty, \tag{3.14}$$

then any sequence $\{\hat{Q}_{ENB}\}$ *of ENB estimators as defined in (3.8), with* $L(n) \to \infty$ *and* $t_n \uparrow \rho$, $[0, t_n] \subset \Theta$, $P-$a.s., *is an AMLE's sequence. Consequently it is strongly convergent, i.e., with probability one,* $\hat{Q}_{ENB} \to Q_0$ *where* Q_0 *is given by H2. Moreover* $\bar{L}_n(\hat{Q}_{ENB}) \to l(Q_0) > -\infty$, $P-$a.s..

Proof: It was shown in Lemma 2.4 *b)* that (3.14) implies $\mathcal{F}_{\mathbf{q}} = \mathcal{P}(\Theta) - \{\varepsilon_0\}$ and therefore *H2* is automatically fullfilled and $l(Q_0) > -\infty$. Moreover $\bar{L}_n(Q_0) \to l(Q_0)$, $P-$a.s.. According to (2.3) it should be proved that, $P-$a.s., $\lim_{n\to\infty} a_n = 0$ where

$$a_n \stackrel{def}{=} \sum_{k\geq 0} p_{n,k}\, ln \left[\frac{p_{\hat{Q}_{MLE}}(k)}{p_{\hat{Q}_{ENB}}(k)}\right].$$

Clearly $\liminf_{n\to\infty} a_n \geq 0$. We define

$$b_n \stackrel{def}{=} \sum_{k\geq 0} p_{n,k}\, ln \left[\frac{p_{\hat{Q}_{MLE}}(k)}{p_{Q_0}(k)}\right], \quad c_n \stackrel{def}{=} \sum_{k\geq 0} p_{n,k}\, ln \left[\frac{p_{Q_0}(k)}{p_{\hat{Q}_{ENB}}(k)}\right]$$

and we have that, $\forall n$, $a_n = b_n + c_n$. As a consequence of the Proposition 2.5 b) or of the Lemma 3.2, it can be deduced that under (3.14), $P-$a.s., $\lim_{n\to\infty} b_n = 0$. Thus $\liminf_{n\to\infty} c_n \geq 0$, $P-$a.s.. In order to close the proof we show that $\limsup_{n\to\infty} c_n \leq 0$, $P-$a.s..

As, $P-$a.s., $L(n) \to \infty$ and $t_n \uparrow \rho$, from Lemma 3.1 there exists $\check{Q}_n \in \mathcal{P}_{L(n),t_n}$ such that $\check{Q}_n \to Q_0$ weakly, $P-$a.s.. From Lemma 3.2 we have that, $P-$a.s., $\lim_{n\to\infty} \bar{L}_n(\check{Q}_n) = l(Q_0) > -\infty$. On the other hand, by the definition of the ENB estimator we have that $\bar{L}_n(\hat{Q}_{ENB}) \geq \bar{L}_n(\check{Q}_n)$ and consequently

$$\liminf_{n\to\infty} \bar{L}_n(\hat{Q}_{ENB}) \geq l(Q_0) > -\infty, \tag{3.15}$$

$P-$a.s.. Now is clear that $\limsup_{n\to\infty} c_n \leq 0$, $P-$a.s..

Q.E.D.

Say that the ENB estimator is $\hat{Q}_{ENB} = \sum_{l=1}^{L(n)} D(u_0 + n\pi_l,\ s_0 + n\pi_l\theta_l b'(\theta_l))$. Then it can be observed (see (3.4)) that

$$\sum_{l=1}^{L(n)} \pi_l \, \frac{s_0 + n\pi_l\theta_l b'(\theta_l)}{u_0 + n\pi_l} = \int_\Theta \theta b'(\theta)\, \hat{Q}_{ENB}(d\theta) = E_{F_{\hat{Q}_{ENB}}}[X_1], \tag{3.16}$$

if $\Theta = \Theta_R$ and $\lim_{\theta\uparrow R} b(\theta) = \infty$.

The ENB estimators are absolutely continuous distributions. If we want to obtain an estimator of the mixing distribution under the form of a mixture of an absolutely continuous distribution with a discrete, finitely supported (with a known number p of support points) distribution we may define

$$\begin{aligned} \tilde{\mathcal{P}}_{L(n),t_n,\pi_n} \ \stackrel{def}{=} \ \{ \, \pi Q + (1-\pi) \sum_{l=1}^{p} p_l \varepsilon_{\theta_l} \ \mid \ Q \in \mathcal{P}_{L(n),t_n}, \ \pi \in [\, \pi_n, 1\,], \\ \theta \in [\, 0, t_n], \ p_l \in [\, 0, 1\,], \ \sum_{l=1}^{p} p_l = 1 \, \} \end{aligned} \tag{3.17}$$

and replace $\mathcal{P}_{L(n),t_n}$ by $\tilde{\mathcal{P}}_{L(n),t_n,\pi_n}$ in the definition (3.8) of the ENB estimator. The role of π_n is to assure that in this case the estimator will not become a discrete distribution. Convergence of such mixed (discrete and absolutely continuous) estimators can be obtained as above.

Finally observe that we may want to define, for other PSD's mixture models than Poisson mixtures, estimators obtained for $D(u, s)$ distributions corresponding to the case $b(\theta) = \theta$. The convergence results may be adapted. As the ENB estimators we propose here are a kind of kernel estimator, the choice of the kernel $D(u, s)$ may not be very important and therefore, for more easy computations, a more tractable kernel may be preferred to a natural conjugate one.

4 Empirical results and concluding remarks

In section 3 we proposed an estimator for the mixing distribution for a class of PSD's mixture models. The estimator has the form of a finite mixture of natural conjugate priors of the PSD's compound family, thus is an absolutely continuous estimator, property which is not shared by the MLE. We called it a ENB estimator and we motivated the definition of this estimator by the nonparametric Bayesian approach for estimating mixtures. Under some conditions, without necessarily asking the well specification of the mixture model, we obtained the convergence of any sequence of ENB estimators.

In this section we study the performances of the ENB estimator compared with the MLE in the case of Poisson mixtures with $\Theta = [0, \infty)$. In this case the estimator we propose has the form of a finite mixture of gamma distributions

$$\hat{Q}_{ENB} = \sum_{l=1}^{L(n)} \pi_l \Gamma(u_0 + n\pi_l,\ s_0 + n\pi_l\theta_l). \tag{4.1}$$

The mixed distribution has the form described in (3.9). Remark that in the case of a good specification of a Poisson mixture model,

$$E_{\mathbf{q}}[\,X_1 ln(X_1+1)\,] < \infty \iff \int_{[\,0,\infty)} \theta\, ln(\theta+1)\, Q_0(d\theta) < \infty \tag{4.2}$$

where Q_0 is the true mixing distribution. In order to justify (4.2) remark that $E_{\mathbf{q}}[\,X_1 ln(X_1+1)\,] = \int_{[\,0,\infty)} f(\theta)\, Q_0(d\theta)$ where $f(\theta) = E_\theta[\,X_1 ln(X_1+1)\,]$ (the expectation E_θ being considered w.r.t. the $\mathrm{Po}\,(\theta)$ distribution). Define $g(\theta) = \sum_{k\geq 0} (1/k!)\, \theta^k\, exp(-\theta)\, ln(k+2)$ and observe that $f(\theta) = \theta\, g(\theta)$ on $[\,0,\infty)$. It can be proved that $h(\theta)/3 \leq g(\theta) \leq h(\theta)$ where $h(\theta) = \int_{[\,0,\theta)} [\,(1-exp(-s))/s\,]\, ds + 3\, ln 2$. As $\lim_{\theta\downarrow 0}[\,h(\theta)/\theta\,] = \lim_{\theta\uparrow\infty}[\,h(\theta)/ln(\theta+1)\,] = 1$, we get (4.2).

We considered sample sizes of, respectively, 25, 50 and 100. For each sample size we generated, independently, 200 i.i.d. samples. We used as mixing distribution the distribution given by the probability density function (p.d.f.) plotted in *Figure 1* with continuous line.

For every sample we performed ENB and ML estimations. For the ENB estimations we used $u_0 = s_0 = 1.5$ which corresponds to a prior guess P_0 (see (3.1)) with the p.d.f. plotted with dashed line in *Figure 1.* The choice of u_0 and s_0 has significant influence only for the shape of some of the distributions involved in the definition of the $\hat{Q}_{ENB}$. One may put more trust in the prior guess G_0 by choosing $w_0 > 0$ for the definition of the ENB estimator (see (3.6)). We took, for any sample, $L(n)$ as being the upper bound for the support size of the MLE, i.e. $L(n) = ([M_n/2]+1) \wedge K(n)$ (see Lemma 2.2).

We plotted in *Figure 2* estimations of the mixing distribution for the generated samples. In every graph, correspondingly, every 40th estimation (dashed lines), the mean of the 200 estimations (continuous lines) and the true p.d. (bold lines) are plotted.

Insert Figure 2 here

For the considered data generating process, in general, the ENB estimator gives a more accurate estimation of the true distribution function (d.f.) for relatively small samples. As it has been expected, when the sample size increases, the ENB estimator gets close to the MLE. We computed the Kolmogorov - Smirnov distances between the estimations and the truth ($d_{K-S}(F,G) = \sup_{-\infty<x<\infty} \mid F(x) - G(x) \mid$).

Insert Figure 3 here

Finally we were interested in the efficiency of the estimations of the probabilities $q_k = P(X_1 = k)$, $k \in \mathbf{N}$. Lambert, Tierney (1984) showed that, if the true mixing distribution has its support in a compact interval $[\,0, M\,]$ and its distribution function does not decrease 'too fast' in any neighborhood of zero, the MLE of q_k is as efficient as the observed probabilities $p_{n,k}$. The results of Lambert, Tierney were extended to PSD's mixtures by Milhaud, Mounime (1995). In *Figure 4* we plotted the smoothed estimations of q_0 given by the ENB estimations (continuous line), the ML estimations (dashed lines) and the p.d.f. of the distribution *Normal*$(q_0,\, n^{-1}q_0(1-q_0))$. We conjecture that the ENB estimator is as efficient as the MLE in estimating the probabilities q_k for PSD's mixture models as considered in this paper.

Insert Figure 4 here

Further work will try to justify the empirical findings about the estimation of the q_k's and to determine and/or to compare the speeds of convergence for the estimations of the distribution function of the mixing distribution obtained, respectively, from the MLE and from a ENB estimator.

References

FLORENS J.P, MOUCHART M., ROLIN J.-M. (1992) Bayesian analysis of mixtures: some results on exact estimability and identification *Bayesian Statistics 4, J. Berger, A. Dawid. and A. Smith, Eds,* 127-145.

LAMBERT D., TIERNEY L. (1984) Asymptotic properties of the maximum likelihood estimates in mixed Poisson model *Ann. Statist.* ***12***, 1388-1399.

LINDSAY B. (1983) The geometry of mixture likelihoods, Part I: A general theory, *Ann. Statist.* ***11***, 86-94.

LÜXMANN-ELLINGHAUS U. (1987) On the identifiability of mixture of infinitely divisible power series distributions, *Statist. and Probab. Letters* ***5***, 375-378.

MACEACHERN S., MÜLLER P. (1994) Estimating Mixtures of Dirichlet Process Model, ISDS Discussion Paper 94-11, Duke University.

MILHAUD X., MOUNIME S. (1995) A Modified Maximum Likelihood Estimator for Infinite Mixtures, *Preprint, Université de Toulouse 3.*

PEARSON, K. (1894) Contributions to the mathematical theory of evolution, *Philos. Trans. Roy. Soc. London Ser. A* ***185***, 71-110.

PFANZAGL J. (1988) Consistency of maximum likelihood estimators for certain non-parametric families, in particular: Mixtures *J. Statist. Plann. Inference* ***19*** , 137-158.

SIMAR L. (1976) Maximum likelihood estimation of compound Poisson process, *Ann. Statist.* ***4***, 1200-1209.

TITTERINGTON D., SMITH A., MAKOV U. (1985) *Statistical analysis of finite mixtures distributions*, New York: John Wiley.

VAN DER VAART A.W., WELLNER J. (1992) Existence and consistency maximum likelihood in upgraded mixture models *J. of Multivariate Anal.* ***37***, 133-145.

A Appendix

Proof of Lemma 2.1

If Θ is a compact interval, the set

$$V \stackrel{def}{=} \{(p_\theta(Y_1), ..., p_\theta(Y_{K(n)})) \mid \theta \in \Theta\} \subset I\!R^{K(n)} \tag{A.3}$$

is also compact, due to the continuity of the applications $\theta \mapsto p_\theta(k)$, $\forall k \in I\!N_0$. We write $co(V)$ for the convex hull of V and it turns out that $co(V) = U$ where $U \stackrel{def}{=} \{(p_Q(Y_1), ..., p_Q(Y_{K(n)})) \mid Q \in \mathcal{P}(\Theta)\}$ (see, for e.g., Lindsay (1983)). As the convex hull of a compact set in $I\!R^{K(n)}$ is compact, U is also a compact set and the maximisation problem

$$\sup_{u \in U} \sum_{j=1}^{K(n)} p_{n,Y_j} \, ln \, u_j \ , \ \ u = (u_1, ..., u_{K(n)}) \tag{A.4}$$

has a unique solution u^0; this solution is necessarily on the boundary of U. By a theorem of Caratheodory there exists $\hat{Q}_n \in \mathcal{P}(\Theta)$ with at most $K(n)$ support points such that $u^0 = (p_{\hat{Q}_n}(Y_1), ..., p_{\hat{Q}_n}(Y_{K(n)}))$

If Θ is not a compact interval we consider $\overline{\Theta}$, the closure of Θ, with the usual topology making it a compact space and with the corresponding Borel σ-field. Remark that $\overline{V} = \{(p_\theta(Y_1), ..., p_\theta(Y_{K(n)})) \mid \theta \in \overline{\Theta}\}$ and $\overline{V}$ is compact. Take $W \stackrel{def}{=} \{\alpha v \mid 0 \leq \alpha \leq 1, \, v \in \overline{V}\}$ and observe that W is also a compact set, as the image of $[0,1] \times \overline{V}$ through the continuous map $(\alpha, v) \mapsto \alpha v$. Extend the definition of $p_Q(\cdot)$ for Q a subprobability measure, define

$$\tilde{U} = \{(p_Q(Y_1), ..., p_Q(Y_{K(n)})) \mid Q \ \textit{subprobability measure on} \ \overline{\Theta}\}$$

and remark that $co(W) \subset \tilde{U}$. For the converse inclusion observe that

$$(p_Q(Y_1), ..., p_Q(Y_{K(n)})) = \int_{\overline{\Theta}} (p_\theta(Y_1), ..., p_\theta(Y_{K(n)})) \, Q(d\theta) \tag{A.5}$$

for Q subprobability measure on $\overline{\Theta}$. As the r.h.s. of the last equality belongs to $\overline{co(W)} = co(W)$ we can conclude that $co(W) = \tilde{U}$ and $\tilde{U}$ is compact. As a consequence the maximisation problem

$$\sup_{u \in \tilde{U}} \sum_{j=1}^{K(n)} p_{n,Y_j} \, ln \, u_j \ , \ \ u = (u_1, ..., u_{K(n)}) \tag{A.6}$$

has a unique solution denoted (by abuse) as before u^0. It remains to show that there exists $\hat{Q}_n \in \mathcal{P}(\Theta)$ such that $(p_{\hat{Q}_n}(Y_1), ..., p_{\hat{Q}_n}(Y_{K(n)}))$ is the solution of (A.6). It easy to see that u^0 is necessarily on the boundary of the $\tilde{U}$ set. As before, using the result due to Caratheodory, there exists $K(n)$ points in W such that $\tilde{u}$ can be represented as a convex combination of these points. Thus there exists $\beta_1, ..., \beta_{K(n)}$ positive numbers with $\sum_{s=1}^{K(n)} \beta_s \in (0,1]$, and $\theta_1, ..., \theta_{K(n)}$ in Θ such that

$$u^0 = \sum_{s=1}^{K(n)} \beta_s \, (p_{\theta_s}(Y_1), ..., p_{\theta_s}(Y_{K(n)})).$$

It turns out that necessarily $\sum_{s=1}^{K(n)} \beta_s = 1$ since $u' \stackrel{def}{=} (\sum_{s=1}^{K(n)} \beta_s)^{-1} u^0$ also belongs to $\tilde{U}$ and u' gives at least the same value for the objective function in (A.6). Now take $\hat{Q}_n \stackrel{def}{=} \sum_{s=1}^{K(n)} \beta_s \varepsilon_{\theta_s} \in \mathcal{P}(\Theta)$ and remark that $u^0 = (p_{\hat{Q}_n}(Y_1), ..., p_{\hat{Q}_n}(Y_{K(n)}))$.

Q.E.D.

Proof of Lemma 2.4

a) Let $m \stackrel{def}{=} \sup_{Q \in \mathcal{F}_{\mathbf{q}}} l(Q) \leq 0$ and consider $\{Q_j\}_{j \geq 1} \subset \mathcal{F}_{\mathbf{q}}$ such that $l(Q_j) \uparrow m$ as $j \to \infty$. There exists a subsequence $\{Q_{j_i}\}_{i \geq 1} \subset \{Q_j\}_{j \geq 1}$ and Q_0 a subprobability on $I\!R$ such that $Q_{j_i} \to Q_0$ vaguely. From Lemma 2.3 we get $p_{Q_{j_i}}(k) \to p_{Q_0}(k)$ for all $k \geq 0$. Fix $L \in I\!N_0$ arbitrarly. As for all j,

$$l(Q_j) < \sum_{k=0}^{L} q_k \, ln \, p_{Q_j}(k)$$

we get

$$m \leq \sum_{k=0}^{L} q_k \, ln \, p_{Q_0}(k).$$

Letting L to increase to infinity we obtain that $m = l(Q_0)$ and given the definition of Θ it turns out that necessarily $Q_0 \in \mathcal{P}(\Theta)$. As $\mathcal{F}_{\mathbf{q}}$ is a convex set, the strict concavity of the logarithmic function and the identifiability of the mixtures assure the uniqueness of the $Q_0 = \arg \sup\limits_{Q \in \mathcal{F}_{\mathbf{q}}} l(Q) = \arg \sup\limits_{Q \in \mathcal{P}(\Theta)} l(Q)$. Q_0 satisfies the inequality in *H2*.

b) Let fix $k \geq 1$ and $Q \in \mathcal{P}(\Theta)$, $Q \neq \varepsilon_0$. Consider $[\alpha, \beta] \subset \Theta - \{0\}$ such that $Q([\alpha, \beta]) > 0$. We get

$$\begin{aligned} ln \, p_Q(k) &\geq ln \int_{[\alpha,\beta]} p_\theta(k) \, Q(d\theta) \\ &= lnQ([\alpha,\beta]) + ln \left[Q([\alpha,\beta])^{-1} \int_{[\alpha,\beta]} p_\theta(k) \, Q(d\theta) \right] \\ &\geq lnQ([\alpha,\beta]) + Q([\alpha,\beta])^{-1} \int_{[\alpha,\beta]} lnp_\theta(k) \, Q(d\theta). \end{aligned} \tag{A.7}$$

where the last inequality is due to Jensen's inequality. A PSD defined by a function $a(\cdot)$ as in *H1* has the property $a_k \geq b_1^k/k!$ since $a^{(k)}(0) = a(0)\{[b'(0)]^k + c_k\}$ with $c_k \geq 0$. Therefore

$$ln\, p_Q(k) \geq lnQ([\alpha, \beta]) + k\, lnb_1 - ln(k!) + k\, ln\alpha - b(\beta). \tag{A.8}$$

It is clear that $\sum_{k\geq 0} q_k(k\, lnb_1 - ln(k!) + k\, ln\,\alpha)$ finite implies $l(Q) > -\infty$. Due to the Stirling's formula, $\forall k \geq 2$, $\exists \lambda_k \in (0,1)$ such that

$$ln(k-1)! = \left(k - \frac{1}{2}\right)\, lnk - k + \frac{1}{2}ln2\pi + \frac{\lambda_k}{12k}. \tag{A.9}$$

Thus $l(Q) > -\infty$ if $\sum_{k\geq 1} q_k k\, ln(k+1) < \infty$, which proves *b)*. Remark that in fact we have proved that, given the condition in (2.9), $\mathcal{F}_{\mathbf{q}} = \mathcal{P}(\Theta) - \{\varepsilon_0\}$.

c) Assume that $\mathcal{F}_q \neq \emptyset$. We have that $\forall\, Q \in \mathcal{P}(\Theta)$, $Q \neq \varepsilon_0$, $\forall k \geq 2$, $\exists \lambda_k \in (0,1)$ such that

$$\begin{aligned} ln\, p_Q(k) &\leq k\, lnk - k - ln(k\,!) + ln\, Q\,[(0,\infty)] \qquad \text{(A.10)} \\ &= -\frac{1}{2}\, lnk - \frac{1}{2}\, ln\, 2\pi - \frac{\lambda_k}{12\, k} + ln\, Q[(0,\infty)]. \end{aligned}$$

The inequality is a consequence of the fact that $\forall k \geq 1$, $\theta^k exp(-\theta) \leq k^k exp(-k)$ and for the equality we used (A.9). From (A.10) we get, for any $Q \in \mathcal{F}_q$

$$-\infty < l(Q) \leq -\frac{1}{2}\sum_{k\geq 2} q_k\, ln\, k - \frac{1}{2}ln\, 2\pi - \sum_{k\geq 2} q_k \frac{\lambda_k}{12\, k}.$$

Necessarily $\sum_{k\geq 0} q_k\, ln(\,k+1) < \infty$.

For the converse implication define, for e.g., $Q(\{k\}) = 6/[\pi(k+1)]^{-2}$, $k \in I\!N_0$. For such a Q, $\forall k \geq 1$,

$$\begin{aligned} ln\, p_Q(k) > &\; ln(6/\pi^2) - 2\, ln(k+1) + k\, ln\, k - k - ln\,(k!) \qquad \text{(A.11)} \\ = &\; ln(6/\pi^2) - 2\, ln(k+1) - \frac{1}{2}\, ln\, k - \frac{1}{2}\, ln\, 2\pi - \frac{\lambda_k}{12\, k}. \end{aligned}$$

From (A.11) it turns out that $Q \in \mathcal{F}_{\mathbf{q}}$, thus $F_{\mathbf{q}}$ is nonempty, if $E_{\mathbf{q}}[\, ln(X_1+1)] < \infty$.

Q.E.D.

Proof of Lemma 3.1

a) The same ideas as for proving $\lim_{r\to\infty} \| f \|_{L^r([a,b])} = \| f \|_{L^\infty([a,b])}$ can be used. For more simple notations we drop the n indices for the rest of the proof. We write

$$
\begin{aligned}
[c_{u,s}]^{-\frac{1}{u}} &= \left[\int_\Theta \left(\theta^{\frac{s-1}{u-1}} exp(-b(\theta))\right)^{u-1} exp(-b(\theta))\, d\theta\right]^{\frac{1}{u}} \\
&\leq \left[\max_{\theta\in\Theta} \theta^{\frac{s-1}{u-1}} exp(-b(\theta))\right]^{\frac{u-1}{u}} \left[\int_\Theta exp(-b(\theta))\, d\theta\right]^{\frac{1}{u}}
\end{aligned}
$$

From the definition of Θ and the properties of $b(\theta)$ it turns out that

$$
\begin{aligned}
\max_{\theta\in\Theta} \theta^{\frac{s-1}{u-1}} exp(-b(\theta)) &\to \max_{\theta\in\Theta} \theta^\gamma exp(-b(\theta)) < \infty \\
\left[\int_\Theta exp(-b(\theta))\, d\theta\right]^{\frac{1}{u}} &\to 1.
\end{aligned}
$$

As a consequence

$$
\limsup_{n\to\infty} [c_{u,s}]^{-\frac{1}{u}} \leq \max_{\theta\in\Theta} \theta^\gamma exp(-b(\theta)).
$$

For a converse inequality let fix $\epsilon > 0$. There exists $N_\epsilon \in I\!N$ and I_ϵ a nondegenerate compact interval included in Θ such that, $\forall n \geq N_\epsilon$,

$$
\theta^{\frac{s-1}{u}} exp(-b(\theta)) \geq \theta^\gamma exp(-b(\theta)) - \epsilon, \quad \forall \theta \in I_\epsilon.
$$

It implies that, for any $n \geq N_\epsilon$,

$$
\begin{aligned}
[c_{u,s}]^{-\frac{1}{u}} &\geq \left[\int_{I_\epsilon} \left(\theta^{\frac{s-1}{u}} exp(-b(\theta))\right)^u d\theta\right]^{\frac{1}{u}} \\
&\geq \left(\max_{\theta\in\Theta} \theta^\gamma exp(-b(\theta)) - \epsilon\right) l(I_\epsilon)^{\frac{1}{u}} \to \max_{\theta\in\Theta} \theta^\gamma exp(-b(\theta)) - \epsilon
\end{aligned}
$$

where $l(I_\epsilon)$ is the length of the interval I_ϵ. As ϵ was arbitrarly fixed we get

$$
\liminf_{n\to\infty} [c_{u,s}]^{-\frac{1}{u}} \geq \max_{\theta\in\Theta} \theta^\gamma exp(-b(\theta)).
$$

b) We show, using *a)* that, $\forall \theta_0 \in \Theta$, $F_{u,s}(\theta_0) \to 0$ if $\theta_0 < \theta(\gamma)$ and $1 - F_{u,s}(\theta_0) \to 0$ if $\theta_0 < \theta(\gamma)$. By $F_{u,s}(\cdot)$ we denote the distribution function of $D(u,s)$. Say that $\theta_0 \in \Theta$ and $\theta_0 < \theta(\gamma)$. It implies that

$$
[F_{u,s}(\theta_0)]^{\frac{1}{u}} = [c_{u,s}]^{\frac{1}{u}} \left[\int_{[0,\theta_0]} \left[\theta^{\frac{s-1}{u}} exp(-b(\theta))\right]^u d\theta\right]^{\frac{1}{u}}.
$$

From the proof of *a)* we get, for large n,

$$
\lim_{s/u\to\gamma} [F_{u,s}(\theta_0)]^{\frac{1}{u}} \leq \frac{\theta_0^\gamma exp(-b(\theta_0))}{\theta(\gamma)^\gamma exp(-b(\theta(\gamma))} < 1.
$$

Necessarily $\lim_{s/u \to \gamma} F_{u,s}(\theta_0) = 0$. Applying the same arguments for $1 - F_{u,s}(\theta_0)$ if $\theta_0 > \theta(\gamma)$, $\theta_0 \in \Theta$, we obtain *b)*.

c) It suffices to prove that any finitely supported p-measure on Θ is a weak limit of a sequence included in $\mathcal{G}$. Consider $\sum_{l=1}^{K} p_l \, \varepsilon_{\theta_l}$ a finite supported p-measure on Θ and say that θ_K is its largest support value. First assume that $\theta_K < \rho$, ρ the upper bound of the interval Θ. As $L(n) \to \infty$ and $t_n \uparrow \rho$, for sufficiently large n, we can build $m_n = \sum_{l=1}^{K} p_l \, D(u_0 + np_l, \, s_0 + np_l \theta_l b'(\theta_l)) \in \mathcal{P}_{L(n),t_n}$. From *b)*, applied for $\gamma = \theta_l b'(\theta_l)$, we have $m_n \to \sum_{l=1}^{K} p_l \, \varepsilon_{\theta_l}$, weakly. If Θ is a compact interval and $\theta_K = \rho$, let say first that $\lim_{\theta \uparrow \rho} b'(\theta) < \infty$. In this case we build m_n with θ_K replaced by t_n and *b)* applies again ($\{t_n\}_{n \geq 1}$ could be in this case a constant sequence equal to ρ). If $\theta_K = \rho$ and $\lim_{\theta \uparrow \rho} b'(\theta) = \infty$ (necessarily $\rho = R$ and $a'(\rho) = \infty$), the density of any $D(u,s)$ attains its maximum in $(0, \rho)$. The argument of this maximum tends to ρ if $s/u \to \infty$. We show that in this case $s/u \to \infty$ implies $F_{u,s} \to \varepsilon_\rho$, weakly. For any $\theta_0 < \rho$, for n large,

$$F_{u,s}(\theta_0) = c_{u,s} \int_{[0,\theta_0\,]} \theta^{s-1} exp(-u \, b(\theta)) d\theta \leq c_{u,s} \left[\theta_0^{\frac{s-1}{u}} exp(-b(\theta_0)) \right]^u \theta_0.$$

From the definition of $c_{u,s}$ and the fact that $c/d \to \infty$, for any $\delta > 0$ such that $\theta_0 + \delta < \rho$, for sufficiently large n,

$$\begin{aligned} c_{u,s}^{-1} &= \int_{[\,0,\,\theta_0+\delta]} \theta^{\,s-1} exp(-u \, b(\theta)) \, d\theta + \int_{[\,\theta_0+\delta,\,\rho\,]} \theta^{\,s-1} exp(-u \, b(\theta)) \, d\theta \\ &\geq \left[(\theta_0 + \delta)^{\frac{s-1}{u}} exp(-b(\theta_0 + \delta)) \right]^u [\, \rho - (\theta_0 + \delta)\,]. \end{aligned}$$

It implies that

$$F_{u,s}(\theta_0) \leq \left[\left(\frac{\theta_0}{\theta_0 + \delta} \right)^{\frac{s-1}{u}} exp(\, b(\theta_0 + \delta) - b(\theta_0)) \right]^u \frac{\theta_0}{\rho - \theta_0 - \delta}$$

and the r.h.s. of the inequality tends to zero. Thus $F_{u,s} \to \varepsilon_\rho$ if $s/u \to \infty$. Finally we may build again m_n with θ_K replaced by $t_n < \rho$, $t_n \uparrow \rho$, and we still have $m_n \to \sum_{l=1}^{K} p_l \, \varepsilon_{\theta_l}$, weakly.

Q.E.D.

Proof of Lemma 3.2

Clearly $l(\check{Q}) > -\infty$. Fix $k \geq 1$ and $Q \in \mathcal{P}(\Theta)$, $Q \neq \varepsilon_0$. Consider $[\,\alpha, \beta\,] \subset \Theta - \{0\}$ such that $Q([\,\alpha, \beta\,]) > 0$. We have from (A.7) and (A.8) that

$$0 \geq ln \, p_Q(k) \geq \alpha_k(Q) \stackrel{def}{=} lnQ([\,\alpha, \beta\,]) + k \, ln(b_1 \alpha) - ln(k!) - b(\beta). \qquad \text{(A.12)}$$

Remark that, $\forall n$, for any $[\,\alpha,\beta\,] \subset \Theta - \{0\}$, $\check{Q}_n(\,[\,\alpha,\beta\,]\,) > 0$ and that

$$\sum_{k\geq 0} p_{n,k}\, \alpha_k(\check{Q}_n) = ln\, \check{Q}_n(\,[\,\alpha,\beta\,])\,) + ln(b_1\alpha) \sum_{k\geq 0} p_{n,k}\, k - \sum_{k\geq 0} p_{n,k}\, ln(k!) - b(\beta) \quad \text{(A.13)}$$

As in Lemma 2.4 *b)* we have that, under the considered hypotheses, $P-$a.s.,

$$\sum_{k\geq 0} p_{n,k}\, ln(k!) \longrightarrow \sum_{k\geq 0} q_k\, ln(k!) < \infty. \quad \text{(A.14)}$$

On the other hand, $P-$a.s.,

$$\sum_{k\geq 0} p_{n,k}\, k \longrightarrow \sum_{k\geq 0} q_k\, k < \infty. \quad \text{(A.15)}$$

Consider $\alpha_k(\check{Q}_n)$ as defined in (A.12) for an interval $[\,\alpha,\beta\,] \subset \Theta - \{0\}$ such that $\check{Q}(\,[\,\alpha,\beta)\,]) > 0$ and $\check{Q}(\{\alpha\}) = \check{Q}(\{\beta\}) = 0$. As a consequence of (A.15) and (A.14) and of the hypothesis $\check{Q}_n \to \check{Q}$ weakly, we have, $P-$a.s.,

$$\sum_{k\geq 0} p_{n,k}\, \alpha_k(\check{Q}_n) \longrightarrow \sum_{k\geq 0} q_k\, \alpha_k(\check{Q}) > -\infty. \quad \text{(A.16)}$$

Let's fix arbitrarly $L \in I\!N_0$. From (A.12) we get that

$$\sum_{k=0}^{L} p_{n,k}\, \left[ln\, p_{\check{Q}_n}(k) - \alpha_k(\check{Q}_n) \right] \leq \sum_{k\geq 0} p_{n,k}\, \left[ln\, p_{\check{Q}_n}(k) - \alpha_k(\check{Q}_n) \right]$$

Letting n to increase to infinity we obtain, $P-$a.s.,

$$\sum_{k=0}^{L} q_k\, ln\, p_{\check{Q}}(k) - \sum_{k=0}^{L} q_k\, ln\, \alpha_k(\check{Q}) \leq \liminf_{n\to\infty} \sum_{k\geq 0} p_{n,k}\, ln\, p_{\check{Q}_n}(k) - \sum_{k\geq 0} q_k\, ln\, \alpha_k(\check{Q}).$$

For $L \to \infty$ we deduce, $P-$a.s.,

$$l(\check{Q}) = \sum_{k\geq 0} q_k\, ln\, p_{\check{Q}}(k) \leq \liminf_{n\to\infty} \sum_{k\geq 0} p_{n,k}\, ln\, p_{\check{Q}_n}(k).$$

From an inequality as (2.18) we get, as in Proposition 2.5 *b)*, $\limsup_{n\to\infty} \bar{L}_n(\check{Q}_n) \leq l(\check{Q})$.

Q.E.D.

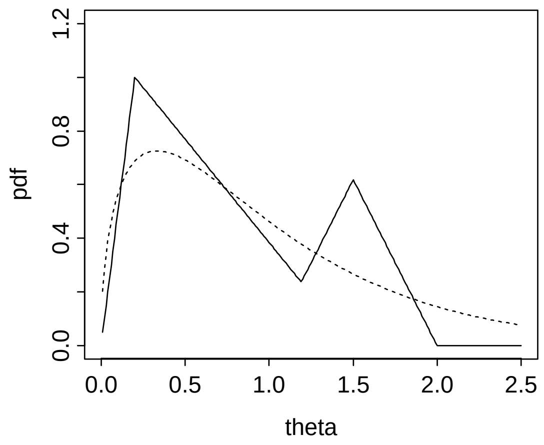
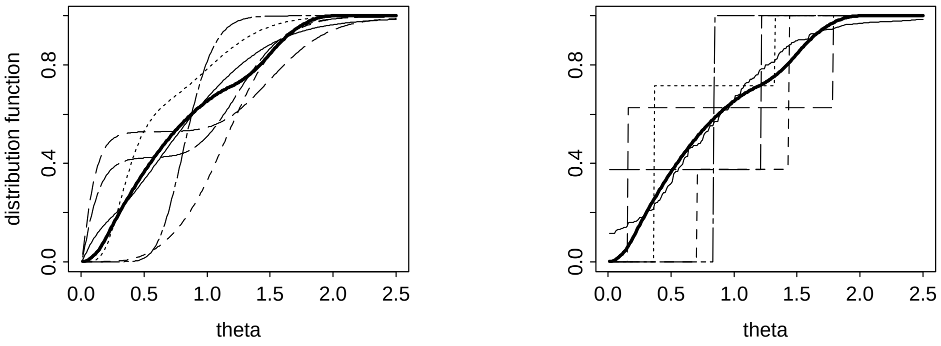
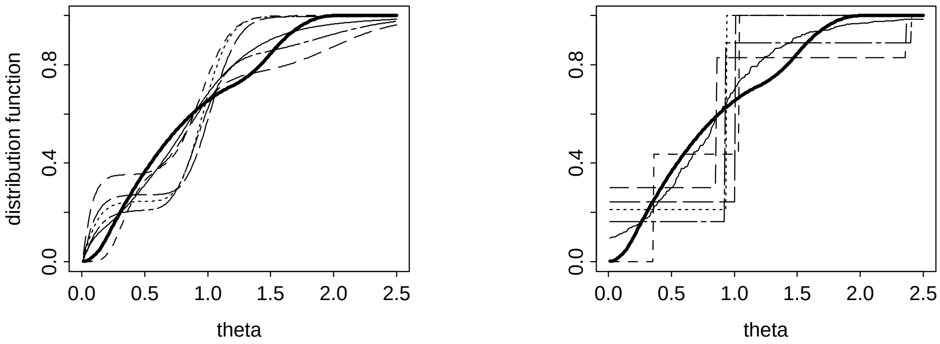


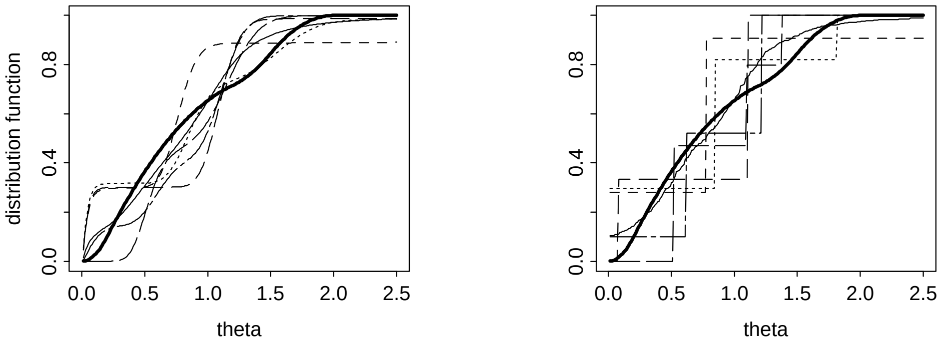
Figure 1.



a) $n = 25$

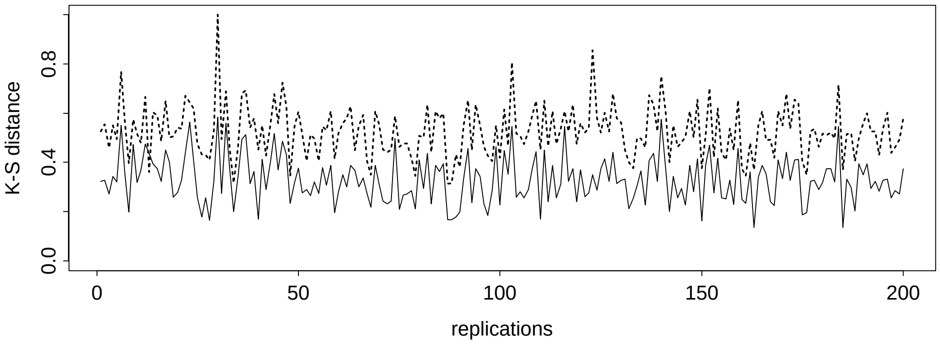


b) $n = 50$

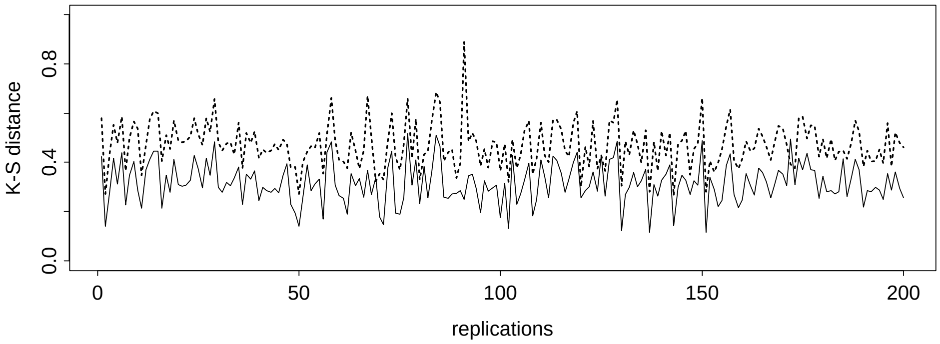


c) n=100

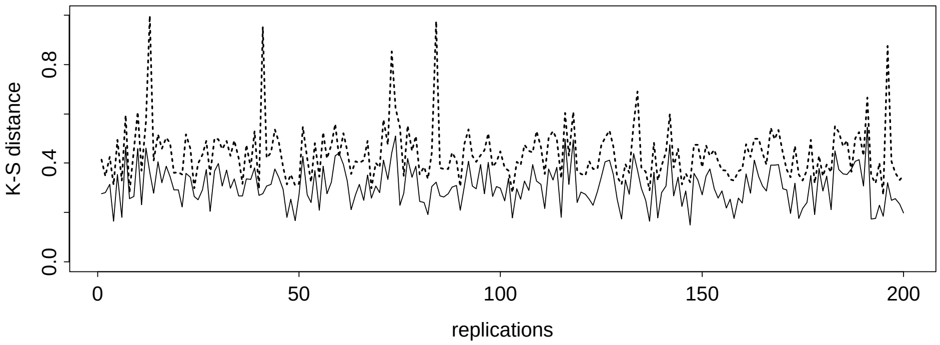
Figure 2 a), b), c). Estimations of the d.f.. Left side - ENB estimations, right side - ML estimations. Dashed lines - estimations, continuous lines - averages of 200 estimations, bold lines - true d.f..



a) n = 25

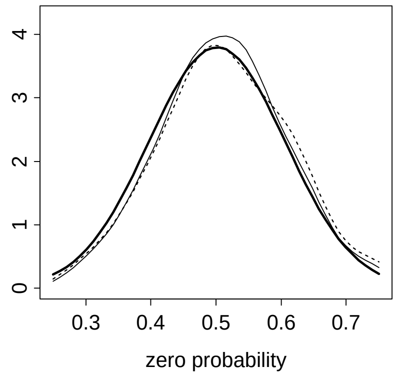


b) n = 50

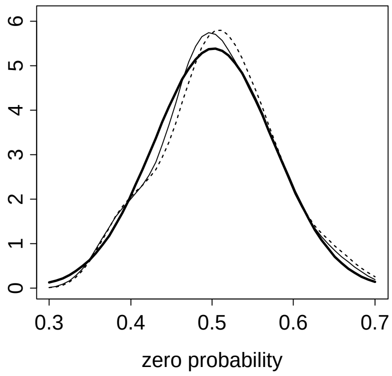


c) n=100

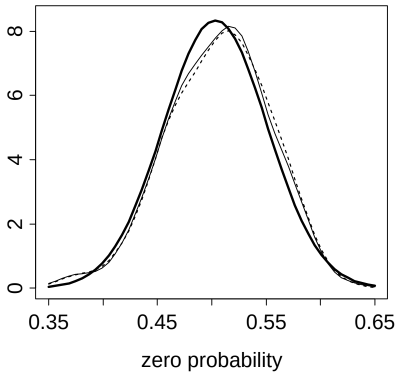
Figure 3 a), b), c). The Kolmogorov-Smirnov distances between the estimated and the true distribution functions computed for 200 samples. The continuous lines corespond to ENB estimators, the dashed lines to MLE's.



a) n=25



b) n=50



c) n=100

Figure 4 a), b), c). Estimations of the zero probability, i.e. $q_0 = P(X_1 = 0)$. *Continuous lines - the zero probabilities given by the ENB estimators, dashed lines - the MLE's of* q_0. *Bold lines - pdf of normal distributions.*