

Quasi-Bayes properties of a sequential procedures for mixtures

Sonia Petrone

Bocconi University, Milan

joint work with **Sandra Fortini**

Statistics Conference in honor of Aad's 60th birthday
Leiden, 17-21 June 2019

Among the many areas to which Aad has given influential contributions

- Semi- and non-parametric inference
- Bayesian nonparametrics
- frequentist properties of Bayesian procedures

Among the many areas to which Aad has given influential contributions

- Semi- and non-parametric inference
- Bayesian nonparametrics
- frequentist properties of Bayesian procedures

and, my master students know (part of:) his lecture notes:

TIME SERIES

A.W. van der Vaart

Universiteit Leiden

version 242013

Among the many areas to which Aad has given influential contributions

- Semi- and non-parametric inference
- Bayesian nonparametrics
- frequentist properties of Bayesian procedures

This talk: a procedure for online nonparametric estimation of the mixing distribution in mixture models

Among the many areas to which Aad has given influential contributions

- Semi- and non-parametric inference
- Bayesian nonparametrics
- frequentist properties of Bayesian procedures

This talk: a procedure for online nonparametric estimation of the mixing distribution in mixture models

but studying its quasi-Bayes properties..



problem and motivation

Premise (predictive approach)

Quasi-Bayes properties of a sequential procedure for mixtures

Asymptotic posterior distribution & credible intervals

discussion

problem and motivation

unsupervised sequential learning and classification

SUMMARY

Coherent Bayes sequential learning and classification procedures are often useless in practice because of ever-increasing computational requirements. On the other hand, computationally feasible procedures may not resemble the coherent solution, nor guarantee consistent learning and classification. In this paper, a particular form of classification problem is considered and a “quasi-Bayes” approximate solution requiring minimal computation is motivated and defined. Convergence properties are established and a numerical illustration provided.

Keywords: SEQUENTIAL; CLASSIFICATION; BAYESIAN; PATTERN RECOGNITION; STOCHASTIC APPROXIMATION; MIXTURES

1. INTRODUCTION

IN this paper, we shall consider a problem of unsupervised sequential learning and classification that has received much attention in the contexts of Pattern Recognition and Signal Detection.

A sequence of (possibly vector-valued) observations $x_1, x_2, \dots, x_n, \dots$, is obtained, each belonging to one of k (≥ 2) exclusive populations $\Pi_1, \Pi_2, \dots, \Pi_k$ (for example, “pattern types” or “signal sources”). For each n , the n th observation is to be classified on the basis of the observed values $x_1, x_2, \dots, x_n$ and without any feedback concerning the correctness, or otherwise, of previous classifications.

We shall consider the situation in which, conditional on $\theta = (\theta_1, \theta_2, \dots, \theta_k)^T$ and density functions $f_1, f_2, \dots, f_k$, we may assume that the random variables $\tilde{x}_n$ are independent, with probability densities

$$f(x_n | \theta) = \theta_1 f_1(x_n) + \theta_2 f_2(x_n) + \dots + \theta_k f_k(x_n), \quad (1)$$

where the θ_i 's are non-negative and sum to unity. The density f_i specifies the probability

J. R. Statist. Soc. B (1978),
40, No. 1, pp. 106–112

A Quasi-Bayes Sequential Procedure for Mixtures

By A. F. M. SMITH† and U. E. MAKOV

University College London

[Received June 1977. Revised November 1977]

SUMMARY

Coherent Bayes sequential learning and classification procedures are often useless in practice because of ever-increasing computational requirements. On the other hand, computationally feasible procedures may not resemble the coherent solution, nor guarantee consistent learning and classification. In this paper, a particular form of classification problem is considered and a “quasi-Bayes” approximate solution requiring minimal computation is motivated and defined. Convergence properties are established and a numerical illustration provided.

Keywords: SEQUENTIAL; CLASSIFICATION; BAYESIAN; PATTERN RECOGNITION; STOCHASTIC APPROXIMATION; MIXTURES

1. INTRODUCTION

In this paper, we shall consider a problem of unsupervised sequential learning and classification that has received much attention in the contexts of Pattern Recognition and Signal Detection.

A sequence of (possibly vector-valued) observations $x_1, x_2, \dots, x_n, \dots$, is obtained, each belonging to one of k (≥ 2) exclusive populations $\Pi_1, \Pi_2, \dots, \Pi_k$ (for example, “pattern types” or “signal sources”). For each n , the n th observation is to be classified on the basis of the observed values $x_1, x_2, \dots, x_n$ and without any feedback concerning the correctness, or otherwise, of previous classifications.

sequential learning on the mixing distribution G

Bayesian DP mixture model

$$\begin{aligned} X_i | \theta_i &\stackrel{\text{ind}}{\sim} f(x | \theta_i) \\ \theta_i | G &\stackrel{\text{iid}}{\sim} G, \text{ with } G \sim \pi \end{aligned}$$

Integrating the θ_i out

$$X_i | G \stackrel{\text{iid}}{\sim} F_G(x) = \int f(x | \theta) dG(\theta)$$

Interest in **sequentially estimating G** , (or the mixing **density g**), as new data arrive.

The Bayesian estimate, w.r.t. quadratic loss, is

$$\hat{G}^{\text{Bayes}}(\cdot) = E(G(\cdot) | x_{1:n}) = P(\theta_{n+1} \in \cdot | x_{1:n})$$

Thus, the problem can be rephrased as **recursive prediction**.

Approximate Bayesian solutions

The Bayesian solution is clear, but computationally demanding, especially with **streaming data and online learning**

In the information- vs computation-efficiency trade-off, **slightly less efficient but more tractable** models are an attractive compromise.

→ Algorithms that can give a fast, although approximate, solution are often designed as **approximation of an optimal, but computationally intractable, Bayesian solution** (e.g. ABC, variational Bayes,...; and recent *sequential* VB,...)

Yet, **what properties?** Understanding how close these approximate solutions are to the Bayesian solution is often difficult. Some results on **frequentist properties** (e.g. consistency of VB)

But, **their Bayesian motivation gets lost..**

A sequential procedure for mixtures

Motivated by computations in Dirichlet process (DP) mixture models, M. Newton et al.¹ proposed a recursive estimate for G that starts at G_0 and for $n \geq 1$ recursively computes

$$\mathbf{G}_n(\mathbf{A}) = (\mathbf{1} - \mathbf{a}_n)\mathbf{G}_{n-1}(\mathbf{A}) + \mathbf{a}_n \mathbf{p}_{\mathbf{G}_{n-1}}(\theta_n \mid \mathbf{x}_n)$$

with $a_i \in (0, 1)$; a *default choice* is $a_n = \frac{1}{\alpha + n}$; or, $\sum_i a_i = \infty$ and $\sum_i a_i^2 < \infty$.

This extends the quasi-Bayes sequential procedure by Smith & Makov (1978) for recursive learning and classification.

¹Newton and Zhang, 1999; Newton, Quintana & Zhang (1998); Newton, 2002

Frequentist consistency

Convergence results have proven the validity of the recursive rule $\hat{G}_n$ as a **consistent frequentist estimator** of G , for some classes of mixture models.

Newton's rule as a stochastic approximation sequence (Smith & Makov (1978), Martin and Ghosh (2008))

Ghosh and Todkar (2006); Tokdar, Martin & Ghosh (*Ann Stat*, 2009)

But the original motivation, providing a fast approximation of a sequential Bayesian procedure, gets lost.. Given the huge popularity of DP and BNP mixture models, one would like to know **if, and, in case, what Bayesian model** one is approximating...

quasi-Bayes properties?

Still, Newton's algorithm is simple and fast, and quite popular. But, is it an approximation of the Bayesian estimate from an exchangeable DP mixture model?

- In fact, it is **not exchangeable**: the estimate depends on the ordering of $x_1, \dots, x_n$!!
- In fact, it is used to estimate the mixing **density**, g . But the DP a.s. selects **discrete** G .

In what sense, then, it is a **quasi-Bayes** procedure?

A new look at Newton's algorithm (and, a new prior for BNP?)

Our key to address the question is to regard the algorithm as a **predictive learning rule**. That is,

$$G_n(\cdot) = P(\theta_{n+1} \in \cdot \mid x_{1:n}), \quad n \geq 1$$

- What is the implied model?
- Is it, and in what sense, an approximation of a Bayesian mixture model? for what prior? Not a DP...perhaps a prior on absolutely continuous G ?
- And, the algorithm only gives a point estimate. A Bayesian procedure would express the uncertainty. Can we provide a posterior distribution?
- Can we give hints for tuning the algorithm, attenuating the dependence of the estimate on the ordering?

Premise (predictive approach)

Predictive approach

* In the *hypothetical* Bayesian (exchangeable) model:

$X_i \mid \theta \stackrel{iid}{\sim} p(x \mid \theta)$, $\theta \sim \pi(\theta)$, which defines the law P of (X_n) , and the predictive distributions

$$P_n(\cdot) \equiv P(X_{n+1} \in \cdot \mid X_{1:n}), \quad n \geq 1, \text{ with } X_1 \sim P_0$$

* In a *predictive approach*, one starts from assumptions on the *predictive learning rule* (P_n) , $n \geq 1$. The sequence (P_n) characterizes the law P of (X_n) (Ionescu-Tulcea).

If such P is *exchangeable*, then, P -a.s.,

$$P_n \Rightarrow \tilde{F} \quad \text{directing random measure}$$

and $X_i \mid \tilde{F} \stackrel{iid}{\sim} \tilde{F}$; the prior on $\tilde{F}$ is the induced law of the limit.

Quasi-Bayes models

Given the predictive rule (P_n) , the induced law P of (X_n) may not be exchangeable. Still, (Aldous, 1983)

if P_n converges, to a random probability measure $\tilde{F}$, then (X_n) is asymptotically exchangeable

$$(X_{n+1}, X_{n+2}, \dots) \xrightarrow{d} (Z_1, Z_2, \dots)$$

where (Z_n) is exchangeable with directing random measure $\tilde{F}$.

Roughly speaking, for large n , $X_n \mid \tilde{F} \stackrel{iid}{\approx} \tilde{F}$.

We say that P is a **quasi-Bayes** approximation of the exchangeable limit law.

Examples

Many predictive characterizations are based on stochastic processes with reinforcement (review: Pemantle, 2007)

* k color Pólya urn: P is exchangeable, and the prior is Dirichlet distribution

* Blackwell & MacQueen Pólya sequences:

$$P_n(\cdot) = \frac{\alpha F_0(\cdot) + \sum_{i=1}^n \delta_{x_i}(\cdot)}{\alpha + n} = \left(1 - \frac{1}{\alpha + n}\right) P_{n-1} + \frac{1}{\alpha + n} \delta_{x_n}$$

P is exchangeable, and $P_n \rightarrow \tilde{F}$ in total variation, P -a.s., where $\tilde{F} \sim DP(\alpha F_0)$. Such $\tilde{F}$ is a.s. **discrete**.

* **Measure-valued Pólya sequences** (Newton et al. (1999), Bandyopadhyay & Thacker (2016), Mailler and Marckert(2017), Janson, 2019)

$$P_n(\cdot) = (1 - a_n) P_{n-1}(\cdot) + a_n R_{x_n}(\cdot)$$

Conditionally identically distributed sequences (c.i.d.)

New priors for BNP? Find $R_{X_n}(\cdot)$ such that

1. (P_n) is a measure valued martingale
2. (X_n) is stationary

and study the limit of P_n .

Thm (Berti, Pratelli, Rigo, Ann.Prob. 2013). If (X_n) is c.i.d., then

$$\tilde{F} \ll \lambda \text{ } P\text{-a.s.} \quad \text{iff} \quad P_n \ll \lambda \text{ and } P_n \rightarrow \tilde{F} \text{ in TV, } P\text{-a.s.}$$

c.i.d. equivalent to (1):

$$X_{n+k} \mid X_{1:n} \stackrel{d}{=} X_{n+1} \mid X_{1:n}, \quad \text{for any } n \geq 0 \text{ and } k \geq 1,$$

Thm. (Kallenberg, 1988) (X_n) exchangeable iff (X_n) is c.i.d. and stationary

We show that *Newton's predictive learning rule* replaces *exchangeability with a c.i.d.* stochastic dependence; in fact, with a more general notion of c.i.d. *w.r.t. a filtration* $(\mathcal{F}_n)$

Def. Berti, Pratelli, Rigo (2004): (X_n) is *c.i.d. w.r.t. a filtration* $(\mathcal{F}_n)$ if the X_i are identically distributed and

$$X_{n+k} \mid \mathcal{F}_n \stackrel{d}{=} X_{n+1} \mid \mathcal{F}_n, \quad k \geq 1, n \geq 1.$$

Newton: $\mathcal{F}_n = \sigma(X_{1:n}, \theta_{1:n})$

Quasi-Bayes properties of a sequential procedure for mixtures

Sequential estimation of the mixing distribution

Recursively estimate the mixing distribution G

$$X_i \mid \theta_i \stackrel{iid}{\sim} f_G(x) = \int k(x \mid \theta) dG(\theta)$$

Hierarchical formulation:

$$X_i \mid (\theta_n) \stackrel{ind}{\sim} k(x \mid \theta_i)$$

$$\theta_i \mid G \stackrel{iid}{\sim} G, \text{ exchangeable}$$

$$G \sim \text{prior}; \text{ e.g., } G \sim \text{DP}(\alpha G_0).$$

Bayesian solution

Bayesian point estimate of G (quadratic loss)

$$\begin{aligned} G_n^{(Bayes)}(\cdot) &= E(G(\cdot) \mid x_{1:n}) = P(\theta_{n+1} \in \cdot \mid x_{1:n}) \\ &= E(P(\theta_{n+1} \in \cdot \mid \theta_{1:n}) \mid x_{1:n}) \\ &= E\left(\frac{\alpha G_0(A) + \sum_{i=1}^n \delta_{\theta_i}}{\alpha + n} \mid x_{1:n}\right) \\ &= \frac{\alpha G_0(A)}{\alpha + n} + \frac{1}{\alpha + n} \sum_{i=1}^{n-1} P(\theta_i \in \cdot \mid x_{1:n}) \\ &= \frac{\alpha + n - 1}{\alpha + n} \frac{\alpha G_0(A) + \sum_{i=1}^{n-1} P(\theta_i \in A \mid x_{1:n})}{\alpha + n - 1} + \frac{1}{\alpha + n} P_{G_{n-1}^{(Bayes)}}(\theta_n \in A \mid x_n) \end{aligned}$$

One would be tempted by a recursive form

$$G_n^{(Bayes)}(A) = \left(1 - \frac{1}{\alpha + n}\right) G_{n-1}^{(Bayes)}(A) + \frac{1}{\alpha + n} P(\theta_n \in A \mid x_{1:n})$$

BUT, the blue term above is NOT $G_{n-1}^{(Bayes)}$, because it incorporates the info in x_n

Recursive estimate

Recursive estimate: start at G_0 and for $n \geq 1$ recursively compute

$$\mathbf{G}_n(\mathbf{A}) = (\mathbf{1} - \mathbf{a}_n)\mathbf{G}_{n-1}(\mathbf{A}) + \mathbf{a}_n\mathbf{P}_{\mathbf{G}_{n-1}}(\theta_n \mid \mathbf{x}_n)$$

with $a_i \in (0, 1)$; a “default choice” is $a_n = \frac{1}{\alpha+n}$; or, $\sum_i a_i = \infty$ and $\sum_i a_i^2 < \infty$.

We recognize a measure-valued Polya urn flavor but the reinforcement measure depends on x_n .

Results. quasi-Bayes properties

We regard Newton's rule as a **predictive learning rule**:

$$\begin{aligned} X_n \mid \theta_n &\overset{\text{ind}}{\sim} k(\cdot \mid \theta_n) \\ \theta_{n+1} \mid X_{1:n} &\sim G_n, \quad n \geq 1, \quad \text{with } \theta_1 \sim G_0. \end{aligned}$$

What model are we actually using?

Results. quasi-Bayes properties

We regard Newton's rule as a **predictive learning rule**:

$$\begin{aligned} X_n \mid \theta_n &\stackrel{\text{ind}}{\sim} k(\cdot \mid \theta_n) \\ \theta_{n+1} \mid X_{1:n} &\sim G_n, \quad n \geq 1, \quad \text{with } \theta_1 \sim G_0. \end{aligned}$$

What model are we actually using?

The (θ_n) , thus the (X_n) , are **no longer exchangeable**. Thus, there is no G such that $\theta_n \mid G \stackrel{iid}{\sim} G$.

Results. quasi-Bayes properties

We regard Newton's rule as a **predictive learning rule**:

$$\begin{aligned} X_n \mid \theta_n &\stackrel{\text{ind}}{\sim} k(\cdot \mid \theta_n) \\ \theta_{n+1} \mid X_{1:n} &\sim G_n, \quad n \geq 1, \quad \text{with } \theta_1 \sim G_0. \end{aligned}$$

What model are we actually using?

The (θ_n) , thus the (X_n) , are **no longer exchangeable**. Thus, there is no G such that $\theta_n \mid G \stackrel{iid}{\sim} G$.

BUT

Results

The (θ_n) , thus the (X_n) , are **no longer exchangeable**. Thus, there is no G such that $\theta_n \mid G \stackrel{iid}{\sim} G$.

BUT

Result such G exists asymptotically: G_n converges to a random distribution G , a.s.

If the mixture is identifiable, the sequence (θ_n) is **asymptotically exchangeable**, with $\theta_n \mid G \stackrel{iid}{\approx} G$ for large n .

Results

The (θ_n) , thus the (X_n) , are **no longer exchangeable**. Thus, there is no G such that $\theta_n \mid G \stackrel{iid}{\sim} G$.

BUT

Result such G exists asymptotically: G_n converges to a random distribution G , a.s.

If the mixture is identifiable, the sequence (θ_n) is **asymptotically exchangeable**, with $\theta_n \mid G \stackrel{iid}{\approx} G$ for large n .

Result the (X_n) are **c.i.d.**. The predictive density converges in TV to a mixture $f_G(x) = \int k(x \mid \theta) dG(\theta)$, and (X_n) is **asymptotically exchangeable**. Informally, for n large,

$$X_i \mid G \stackrel{iid}{\approx} f_G(x) = \int k(x \mid \theta) dG(\theta).$$

In this sense, Newton's rule is a **quasi-Bayes mixture model**, with prior distribution on G induced by the law P of (X_n) .

- * In some problems, stationarity may actually be broken by forms of selection, competition, or by interventions whose effect tends to vanish. (cfr following example)
- * Or, it's just a convenient assumption for computational tractability. Then we do want, at least, an asymptotic exchangeable mixture model!
- * Notice that G is in general different from the original prior (e.g. DP): Newton's model is changing the prior.

Thm *Under fairly mild conditions, if G_0 is abs continuous then G is abs continuous, P -a.s.*

Example: non-stationary mixture

Suppose data are collected **over time**, after a campaign that has an *unpredictable* effect on the population, that tends to vanish.

Then, **the θ_i are not exchangeable**: order (time) matters. Thus, their probability law cannot be represented as $\theta_i \mid G \stackrel{iid}{\sim} G$. Rather, one could think of **a sequence $\tilde{G}_n$** :

$$\theta_n \mid \tilde{G}_n \stackrel{ind}{\sim} \tilde{G}_n.$$

Unpredictable: the conditional law of the new $\tilde{G}_n$ is a DP centered on the current estimate G_{n-1}

$$\tilde{G}_n \mid X_{1:n-1}, \theta_{1:n-1} \sim DP((\alpha + n - 1)G_{n-1})$$

In this case, **Newton's one-step-ahead update is THE Bayesian estimate** of $\tilde{G}_n$ from the DP prior.

What loss of information?

For DP mixture models,

$$G_n^{Bayes}(\cdot) = P(\theta_{n+1} \in \cdot \mid x_{1:n}) = \frac{\alpha G_0(\cdot) + \sum_{i=1}^n P(\theta_i \in \cdot \mid \mathbf{x}_{1:n})}{\alpha + n}$$

Regarding Newton's rule as a **predictive learning rule**, we can show that

$$\mathbf{G}_n(A) = P(\theta_{n+1} \in \cdot \mid x_{1:n}) = \frac{\alpha G_0(A) + \sum_{i=1}^n P(\theta_i \in A \mid \mathbf{x}_{1:n})}{\alpha + n}$$

Intuitively, the loss of information is small if the loss due to the latent nature of the θ_n is small.

Simulation study: prior on G

Location mixture of Gaussians, σ^2 known.

If σ^2 is small, G has a prior close to a DP (but abs. continuous!)

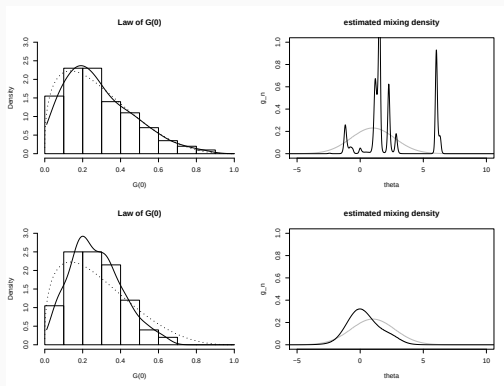


Figure 1: Prior on $G(0)$ (left; dotted: DP (Beta)) and estimated mixing density (right). $\sigma^2 = .01$ (first row) and $\sigma^2 = 1$ (second row)

Asymptotic posterior distribution & credible intervals

Asymptotic posterior distribution

Although the prior is only implicitly defined, we can give results on the induced asymptotic posterior distribution.

Indeed, a Bayesian mixture model would offer more than the point estimate provided by the algorithm.

To be *quasi-Bayes*, we want a full **posterior distribution of $G \mid x_{1:n}$** .

We provide an a.s. asymptotic Gaussian approximation of the posterior law of $(G(A_1), \dots, G(A_k))$, given $x_{1:n}$. For brevity, here $k = 1$.

For any measurable set A , let

$$V(A) = \int_{\{x: f_G(x) \neq 0\}} P_G(A \mid x)^2 dF_G(x) - G(A)^2$$

and consider

$$V_{n,A} = \int P_{G_n}(A \mid x)^2 dF_{G_n}(x) - G_n(A)^2$$

Lemma: For any measurable A , $V_n(A) \rightarrow V(A)$ a.s.

Results

Thm. Let (a_n) satisfy $\sum_n a_n = \infty$ and $\sum_n a_n^2 < \infty$ and let (r_n) be a monotone sequence of positive numbers such that $r_n \sim (\sum_{k>n} a_k^2)^{-1}$ as $n \rightarrow \infty$. If

$$\sqrt{r_n} \sup_{k \geq n} \alpha_k \rightarrow 0 \text{ and } \sum_{k \geq 1} r_k^2 \alpha_{k+1}^4 < \infty,$$

then, for every measurable A ,

$$P(\sqrt{r_n}(G(A) - G_n(A)) \leq t \mid X_{1:n}) \rightarrow \Phi(t \mid 0, V_A).$$

If $\alpha_n = (\alpha + n)^{-\beta}$ with $1/2 < \beta \leq 1$ and $\alpha > 0$, then the above holds with $r_n = (2\beta - 1)n^{2\beta-1}$.

Remark: $V_A = 0$ if and if $G(A)$ is either zero or one.

Results

In the previous result, the limit variance V_A is unknown, depending on G . Replacing the random V_A with the convergent estimate $V_{n,A}$ is not so direct as using Cramér-Slutsky Theorem in standard i.i.d. settings; but we get

Theorem Let $(\theta_n, X_n) \sim P$ and $G = \lim G_n$. Then, under the same assumptions as before, for almost all $\omega = (x_1, x_2, \dots)$ such that $V_A(\omega) > 0$,

$$P(\sqrt{r_n} \frac{G(A) - G_n(A)}{\sqrt{V_{A,n}}} \leq t \mid x_{1:n}) \rightarrow \Phi(t \mid 0, 1).$$

Ex: $\alpha_n = 1/(\alpha + n)$, then $r_n = n$.

This allows to obtain **asymptotic credible intervals** for $G(A)$, given $x_{1:n}$. If $V_A \neq 0$,

$$P(G_n(A) - z_{1-\gamma/2} \sqrt{V_{n,A}/n} < \mathbf{G(A)} < G_n(A) + z_{1-\gamma/2} \sqrt{V_{n,A}/n} \mid \mathbf{x}_{1:n}) \approx 1 - \gamma$$

Remark 1. These results differ from Bernstein-von-Mises types of theorems, which are states a.s.- $P_{G_{true}}^\infty$. They inform about the rate of convergence of $G_n(A) = G_n(\cdot)(x_{1:n})$ to the limit $G(A)(x_1, x_2, \dots)$, a.s. P .

Remark 2. The lack of exchangeability implies that **the estimates obtained by Newton's algorithm depend on the ordering of the observations**. Newton et al, Todkar, Martin, Ghosh (2009): use an average over permutations of the sample.

→ Having better understood what is the probabilistic model underlying Newton's learning rule gives practical insights on **how tuning the algorithm to attenuate the effect of the ordering**

In the following plots

$$\alpha_n = \begin{cases} 1/(\alpha + n) & \text{for } n \leq 500 \\ 1/(\alpha + n)^{3/4} & \text{for } n > 500 \end{cases}$$

with $\alpha = 100$.

Tuning the weights to attenuate the effect of the order

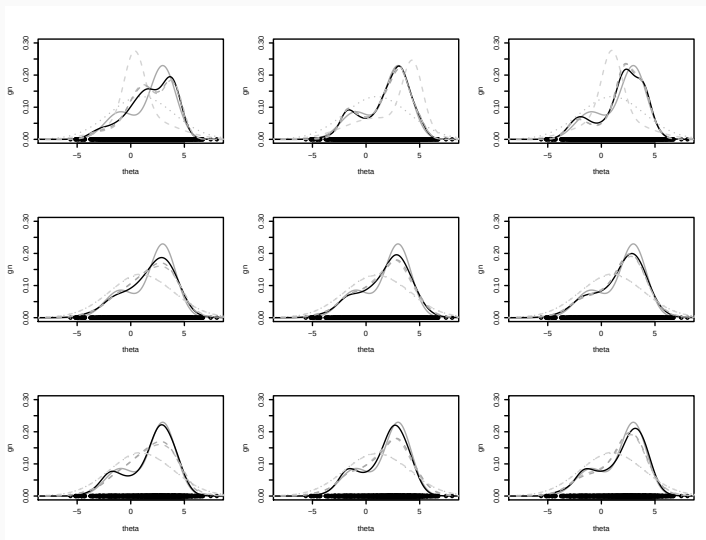


Figure 2: Newton's recursive estimate of the mixing density (black). The data in the first column are a simulated random sample of size 1000 from a location mixture

(marginal) credible intervals for G

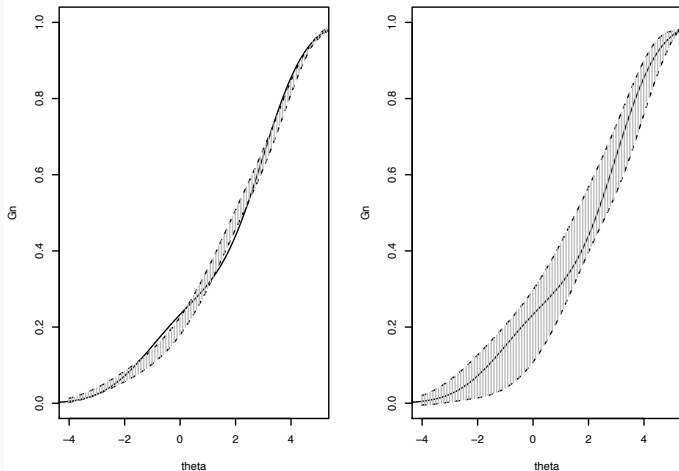


Figure 3: Recursive estimate of the mixing distribution: marginal asymptotic 0.95 credible intervals for $G(t_j)$, for t_j as evidenced in the plot. The black line is the true

discussion

- Newton et al.'s recursive algorithm requires to **evaluate an integral** at each step.

A clever recursive algorithm, that avoids this difficulty: Hahn, Martin, Walker (2018).

- Developments for **multivariate mixtures**, and **dependent mixture models**: Cappello & Walker (2018).

Extensions of our results to dependent mixtures might exploit the notion of families of **partially c.i.d.** sequences (Fortini & P. & Sporysheva, 2017).

- ...

thank you, and... **buon compleanno, Aad**,