ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Convergence of an iterative algorithm to the nonparametric MLE of a mixing distribution

Minwoo Chae [a,*], Ryan Martin [b], Stephen G. Walker [c]

[a] Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, United States
[b] Department of Statistics, North Carolina State University, United States
[c] Department of Mathematics, University of Texas at Austin, United States

ARTICLE INFO

Article history:
Received 17 November 2017
Received in revised form 12 March 2018
Accepted 4 May 2018
Available online xxxx

MSC:
62G05
62G20

Keywords:
Bayesian update
Deconvolution
Mixture model
Predictive recursion
Smoothing

ABSTRACT

An iterative algorithm has been conjectured to converge to the nonparametric MLE of the mixing distribution. We give a rigorous proof of this conjecture and discuss the use of this algorithm for producing smooth mixing densities as near-MLEs.

© 2018 Elsevier B.V. All rights reserved.

UNCORRECTED PROOF

1. Introduction

Consider a mixture model with density function $f = f_P$ given by

$$f(y) = \int k(y \mid x)\, P(dx) \tag{1}$$

where $k(y \mid x)$ is a known kernel density and P is an unknown mixing distribution. The goal is estimation of P based on independent and identically distributed data $Y_1, \ldots, Y_n$ from the mixture f in (1), a classically challenging problem in statistics. If P is a discrete distribution with fixed and finite number of components, then (1) is a finite mixture model and is relatively straightforward; indeed, maximum likelihood computation is feasible with the EM algorithm (Dempster et al., 1977) and the usual asymptotic theory is available (Redner and Walker, 1984). The catch is that the number of mixture components can be difficult to specify. Therefore, there has also been a lot of work on finite mixtures with an unknown number of components (e.g., Woo and Sriram, 2006; Miller and Harrison, 2017).

Likelihood-based methods for estimating P are available even without explicitly making the problem finite-dimensional. Indeed, the likelihood function for P in the nonparametric case is

$$L(P) = \prod_{i=1}^{n} \int k(Y_i \mid x)\, P(dx), \tag{2}$$

* Correspondence to: 10900 Euclid Avenue - Yost Hall Room 231 Cleveland, OH 44106-7058, USA.
E-mail addresses: minwoo.chae@case.edu (M. Chae), rgmarti3@ncsu.edu (R. Martin), s.g.walker@math.utexas.edu (S.G. Walker).

https://doi.org/10.1016/j.spl.2018.05.012
0167-7152/© 2018 Elsevier B.V. All rights reserved.

and it is known that the maximizer $\widehat{P}$, the nonparametric maximum likelihood estimator (MLE), is discrete, with at most n components (Lindsay, 1983, 1995). Discreteness simplifies computation, and fast algorithms are available, e.g., Wang (2007) and Koenker and Mizera (2014). However, if P is believed to have a density with respect to, say, the Lebesgue measure, then the discrete estimator may not be satisfactory. For example, in the image reconstruction of positron emission tomography (Vardi et al., 1985), the nonparametric MLE often provides unsatisfactory reconstructions. Various smoothed versions of $\widehat{P}$ have been proposed (e.g., Zhang, 1990; Green, 1990; Silverman et al., 1990; Eggermont and LaRiccia, 1995, 1997; Goutis, 1997; Andersen and Hansen, 2001; Liu et al., 2009; Belomestny and Schoenmakers, 2014; Comte and Genon-Catalot, 2015; Rebafka and Roueff, 2015), but some of these are rather complicated and there seems to be no general consensus that one smoothing method is any better than another.

For Bayesian mixture models, the Dirichlet process prior (Ferguson, 1973) and variants of its stick-breaking representation (Sethuraman, 1994) have become a mainstay, largely because of the plethora of powerful Markov chain Monte Carlo methods available for evaluating the corresponding posterior (e.g., Escobar and West, 1995; Walker, 2007). The focus of these developments, however, has been the mixture density, with the mixing distribution serving merely as a modeling tool; but see Nguyen (2013). As with the nonparametric MLE, the inherent discreteness of stick-breaking priors, while advantageous for mixture density estimation and modeling latent structures, is problematic in the context of nonparametric Bayesian estimation of a mixing density.

A Bayesian-style recursive estimate for P, called *predictive recursion*, was proposed by Newton (2002) and studied theoretically by Martin and Ghosh (2008), Martin and Tokdar (2009), and Tokdar et al. (2009). The algorithm is fast and provides an estimator having a smooth density with respect to any specified dominating measure. However, its dependence on the (arbitrary) order in which the data $Y_1, \ldots, Y_n$ is processed, which makes it non-Bayesian, along with its inability to be characterized as an optimizer of any objective function, makes the predictive recursion estimator difficult to interpret.

In this paper, we investigate properties of a simple and fast iterative algorithm, one that shares certain features with the MLE, a Bayesian approach, as well as predictive recursion. Versions of this algorithm have been presented in the literature before, and its convergence properties have been conjectured but not rigorously proved. Here we fill this gap by providing a proof that the algorithm converges to the nonparametric MLE as the number of iterations approaches infinity. While the limit is a discrete distribution, it is interesting that at every finite number of iterations, the algorithm returns a continuous density. This suggests that a smooth *near-MLE* of the density can be obtained by stopping the algorithm before convergence is achieved. In the online supplement, we propose a data-driven stopping rule and demonstrate empirically the quality performance of this nonparametric near-MLE of the mixing density compared to predictive recursion.

2. A simple and fast algorithm

2.1. Definition

Let p be a density of the mixing distribution P in (1) with respect to Lebesgue measure. Given a prior guess p_0 of p, if a data point Y_i is observed, then the Bayesian update of p_0 to $p_{1,i}$, say, is

$$p_{1,i}(x) = \frac{k(Y_i \mid x) p_0(x)}{f_0(Y_i)}, \tag{3}$$

where $f_0(y) = \int k(y \mid x) p_0(x)\, dx$. However, we can carry out this single-observation update for any $i = 1, \ldots, n$ and, since observations ordering is irrelevant, it is reasonable to take an average:

$$p_1(x) = \frac{1}{n} \sum_{i=1}^{n} p_{1,i}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{k(Y_i \mid x) p_0(x)}{f_0(Y_i)}.$$

This same argument can be applied, with p_0 replaced by p_1, to get an updated estimate p_2, and so on. This suggests the following iterative algorithm for an estimator of p:

$$p_{t+1}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{k(Y_i \mid x) p_t(x)}{f_t(Y_i)}, \quad t \geq 0, \tag{4}$$

where $f_t(y) = \int k(y \mid x) p_t(x)\, dx$ for general $t \geq 0$. Algorithms similar to (4) for certain applications or models have appeared in the literature; see, e.g., Vardi et al. (1985), Laird and Louis (1991), and Vardi and Lee (1993). But despite the hints in these papers about more general versions, it seems that the algorithm (4) has not been studied thoroughly and in the level of generality considered here.

Aside from this Bayesian-motivated formulation, there are a number of ways to think about this algorithm and understand what it is trying to do. First, note the similarities with the predictive recursion algorithm of Newton (2002) which updates by taking a weighted average of the current guess and the single-observation Bayes update (3) based on the current guess as the prior. These computations proceed along the sequence $i = 1, \ldots, n$ and, therefore, the result depends on the arbitrary order of the data sequence. The algorithm (4) can, therefore, be viewed as an order-invariant version of predictive recursion that can be refined *ad infinitum*, by taking $t \to \infty$.

UNCORRECTED PROOF

Please cite this article in press as: Chae M., et al., Convergence of an iterative algorithm to the nonparametric MLE of a mixing distribution. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.05.012.

Download English Version:

https://daneshyari.com/en/article/7548043

Download Persian Version:

https://daneshyari.com/article/7548043

Daneshyari.com