MINIMUM DISTANCE ESTIMATION OF MIXTURE MODELS: AN R VINAIGRETTE

ROGER KOENKER

Abstract. Nonparametric maximum likelihood estimation of mixture models has proven to be a valuable tool for empirical Bayes decision making. The question naturally arises: Is there something special about the likelihood criteria, or could other minimum distance criteria serve equally well? If so, can we say something about under what circumstances various criteria perform well. To initiate such an investigation an implementation of a class of Rényi entropy criteria that incorporates most of the usual minumum distance candidates is implemented and some preliminary testing is undertaken.

1. Introduction

In the R package **REBayes**, Koenker and Gu (2015–2024) we have provided functions for solving a wide variety of nonparametric maximum likelihood problems of the form,

$$\min_{G \in \mathcal{G}} \Big\{ n^{-1} \sum_{i=1}^{n} \log f(y_i) \mid f(y) = \int \varphi(y|\theta) dG(\theta). \Big\}$$

In practice, we have found it more efficient to solve the corresponding dual problem,

$$\max_{\nu \in \mathbb{R}^n} \Big\{ n^{-1} \sum_{i=1}^{n} \log \nu_i \mid n^{-1} \sum_{i=1}^{n} \nu_i \varphi(y_i|\theta) \leq n \text{ for all } \theta \Big\},$$

to recover the the primal solution from the dual solution we must solve,

$$n^{-1} \sum_{j=1}^{m} \varphi(y_i|\hat{\theta}_j) g_j = \frac{1}{\hat{\nu}_i}.$$

The dual constraint obviously has to be evaluated on a grid, $\{t_1, \ldots, t_m\}$ of θ's, and we denote the discrete mass of $\hat{G}$ at these points by the g_j's. There are typically $\mathcal{O}(\log n)$ of these corresponding to the active constraints. See Koenker and Mizera (2014), and Polyanskiy and Wu (2020).

We have argued elsewhere, Koenker and Mizera (2010, 2018) that the family of Rényi divergences offers a convenient class of alternatives to the maximum likelihood criterion for shape constrained density estimation. In that context, modifying the fitting criterion to match a concavity constraint was essential to preserve the convexity of the underlying variational problem. For mixture problems there is no such obligation, but several authors have entertained special cases of minimum distance fitting that fall nicely into the Rényi class. Since the implementation of the Rényi alternatives was already available in the `medde` function of the **REBayes** package, it was relatively simple to adapt this code for fitting mixture models. For NPMLE fitting we rely on the function `KWDual` implementation of the dual form of the

August 1, 2024. A genre manifesto for R Vinaigrettes is available at `http://davoidofmeaning.blogspot.com/2016/12/r-vinaigrettes.html`.

	α = -1	α = -0.5	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
n= 500	2.22	2.23	2.21	2.22	2.20	2.94
n= 1000	2.20	2.20	2.19	2.20	2.19	2.93

Table 1. RMSE for several alternative minimum distance estimators: Poisson case

	α = -1	α = -0.5	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
n= 500	0.96	0.97	0.96	0.96	0.98	
n= 1000	0.96	0.97	0.95	0.96	0.98	

Table 2. RMSE for several alternative minimum distance estimators: Gaussian case

discretized probllem. A modified version of this function allows the user to specify a parameter, `alpha`, that selects a fitting criterion from the Rényi menu. Calls to various special fitting functions: `Gmix`, `Pmix`, etc. can pass the the `alpha` parameter to the `KWDual` through the `R` `dots` mechanism. Further details on the implementation of the Rényi fitting criteria in Mosek, ApS (2022), may be found in Koenker (2019).

2. Some Simulation Evidence

To begin to explore performance of these alternative minimum distance fitting criteria, we will consider two canonical empirical Bayes compound decision models: a Poisson mixture model,

$$Y_i \sim \text{Pois}(\theta_i), \quad \theta_i \sim U[0.5, 15], \quad i = 1, \dots, n,$$

and the Gaussian sequence model,

$$Y_i \sim \mathcal{N}(\theta_i, 1), \quad \theta_i \sim U[0.5, 15], \quad i = 1, \dots, n.$$

We consider six values of the Rényi parameter α: $\{-1, -0.5, 0, 0.5, 1, 2\}$. The value $\alpha = 0$ corresponds to maximum likelihood fitting, $\alpha = 0.5$ to Hellinger fitting, $\alpha = 1$ to Shannon entropy, and $\alpha = 2$ to Pearson fitting. Performance is evaluated by a root mean squared error criterion: for each replication of the experiment, we compute,

$$L(\hat{\theta}, \theta) = \Big\{ n^{-1} \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2 \Big\},$$

where $\hat{\theta}_i$ denotes the posterior mean $\mathbb{E}_{\hat{G}}(\theta | Y = y_i)$. These quantities are then averaged over replications of the experiment, and the square root of these averages are reported. Two sample sizes, $n \in \{500, 1000\}$, are considered and the number of replications is set to $R = 1000$.

In Table 1 we report root mean squared errors for the Poisson experiment. The NPMLE procedure, $\alpha = 0$ performs well as does the Shannon entropy procedure, $\alpha = 1$, while the other procedures are very slightly worse, only the Pearson $\alpha = 2$ procedure is clearly inferior.

In Table 2 we report root mean squared errors for the Gaussian experiment. Again, the NPMLE procedure, $\alpha = 0$ performs well as does the Shannon entropy procedure, $\alpha = 1$, while the other procedures do almost as well. However, the Pearson, $\alpha = 2$, setting for the Gaussian model produces unbounded solutions that fail to yield sensible estimates of G. This is rather mysterious and is left for future investigation.

References

ApS, M. (2022), Mosek modeling cookbook, release 3.3.0. Available from `https://docs.mosek.com/MOSEKModelingCookbook-a4paper.pdf`.

Koenker, R. (2019), Notes on the implementation of Rényi penalized density estimation. Available from `http://www.econ.uiuc.edu/~roger/research/vinaigrettes/Mosek9.pdf`.

Koenker, R. and Gu, J. (2015–2024), REBayes: An R package for empirical Bayes methods. Available from `https://cran.r-project.org/package=REBayes`.

Koenker, R. and Mizera, I. (2010), 'Quasi-concave density estimation', *Annals of Statistics* **38**, 2998–3027.

Koenker, R. and Mizera, I. (2014), 'Convex optimization, shape constraints, compound decisions, and empirical Bayes rules', *Journal of the American Statistical Association* **109**, 674–685.

Koenker, R. and Mizera, I. (2018), 'Shape Constrained Density Estimation Via Penalized Rényi Divergence', *Statistical Science* **33**, 510 – 526.

Polyanskiy, Y. and Wu, Y. (2020), Self-regularizing property of nonparametric maximum likelihood estimator in mixture models. Available from `https://arxiv.org/abs/2008.08244`.