

## Chapter 7: Summary and Future Directions

### 7.1 Summary

In this dissertation, we focus on the density estimation problem in an exponential family induced by a RKHS,  $\mathcal{Q}_{\text{ker}}$ . We proposed a new early stopping SM density estimator obtained via minimizing the (unpenalized) SM loss functional by the gradient descent algorithm and terminating early. We studied its statistical properties and compared it with the penalized SM density estimator in the literature and showed their similarities and differences.

We also compared these two kinds of regularized SM density estimators with the penalized ML density estimator. Via numerical examples, we observed that the regularized SM density estimators are very sensitive to the presence of an isolated observation, especially when there is a small amount of regularization, but the penalized ML density estimator does *not*. We attempted to explain why this happens.

In order to understand this phenomenon, we extended the classic notion of the influence function to allow its input to be a function-valued statistical functional. Using this extended influence function, we studied the sensitivity of regularized SM and ML density estimators in both finite-dimensional and kernel exponential families.

As we have mentioned at the end of the preceding chapter, we recommend to use the regularized SM density estimators, mainly due to its computational advantages.

But when using them, one needs to make sure an appropriate amount of regularization has been imposed.

## 7.2 Future Directions

Numerical examples shown in this dissertation are restricted to  $d = 1$ . We conjecture that regularized SM density estimators in higher dimensions ( $d \geq 2$ ) are also very sensitive to the presence of an isolated observation when the regularization is small, and need numerical examples to confirm this.

Several versions of generalized SM loss functionals have been proposed in the literature in recent years. Parry, Dawid, and Lauritzen (2012) proposed a version that does not only involve the first two derivatives of a log-density function but also the higher-order derivatives. Yu, Drton, and Shojaie (2020) proposed the following generalized H-divergence between  $p_0$  and  $q$

$$\int_{\mathcal{X}} p_0(x) \|\nabla \log p_0(x) \odot \sqrt{h(x)} - \nabla \log q(x) \odot \sqrt{h(x)}\|_2^2 dx, \quad (7.1)$$

where  $\odot$  denotes element-wise multiplication of two vectors,  $h : \mathcal{X} \rightarrow [0, \infty)^d$ , and the square root is taken element-wise. Under certain regularity conditions, one can apply integration by parts and obtain a loss functional from (7.1) that depends on  $q$  only. It is interesting to apply these generalized SM loss functionals to the density estimation problem in  $\mathcal{Q}_{\text{ker}}$  and see whether the resulting density estimators are sensitive to the presence of an isolated observation. We conjecture, with an appropriate choice of  $h$  in (7.1), the resulting density estimators may *not* be very sensitive to isolated observations.

As we have discussed in Chapter 1, there has been an increasing interest in log-concave density estimation recently. The dominant approach is to minimize the NLL loss functional over the class of log-concave density functions over  $\mathbb{R}^d$ , resulting in

the ML log-concave density estimator. This approach leads to a non-differentiable objective function to optimize and involves approximations of the normalizing constant and its sub-gradient, which is computationally challenging. In addition, as is shown in Figure 7.1, the ML log-concave density estimates contain ridges, and are only supported over the convex hull of data with all boundary points being the discontinuous points of the density estimate, which are undesirable qualitative features and may cause series issues in statistical applications.

Since the SM loss functional does not involve the normalizing constant and involves the first two partial derivatives of the log-density function, it is interesting to estimate a log-concave density function by minimizing the SM loss functional and compare the resulting SM log-concave density estimator with the ML log-concave density estimator. We conjecture the issues with the ML log-concave density estimator described above can be solved by using the SM loss functional.

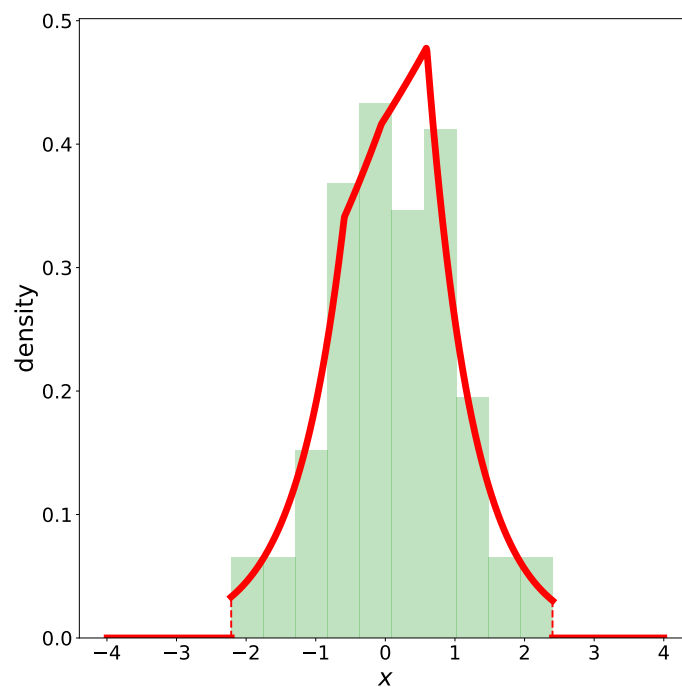


Figure 7.1: ML log-concave density estimate with 100 random samples from the standard normal distribution, where the density estimate is computed using the R package `logcondens` (Dümbgen and Rufibach, 2010). Histogram with the bin width chosen by the Freedman-Diaconis rule is shown in green.