Score-matching methods for log-concave distributions

Dr. Oksana Chernova
joint work with Prof. Mathias Drton
TU Münich

11.10.2022

Shape-constrained estimation

Aim: estimate an unknown density $f : \mathbb{R}^d \to [0, \infty) \in F_d$ based on observations $X_1, ..., X_n$ i.i.d. $\sim f$

Nonparametric inference under shape constraints dates back to [Grenander, 1956], who studied the nonparametric maximum likelihood estimator of a decreasing density on the non-negative half line.

Motivation:

- capture prior information about the problem under investigation
- avoid the need for tuning parameter selection

Properties of log-concave class

- contains many common parametric distributions

  (e.g., normal, gamma with shape ≥ 1, Weibull with exponent ≥ 1, beta with both parameters ≥ 1)
- class is closed under convolution, affine transformation, marginalization
- convergence in distribution

It is a natural infinite-dimensional generalization of the class of Gaussian densities.

Quite flexible class for addressing key contemporary data challenges such as multivariate density estimation and regression.

Properties of log-concave class

- contains many common parametric distributions

  (e.g., normal, gamma with shape ≥ 1, Weibull with exponent ≥ 1, beta with both parameters ≥ 1)
- class is closed under convolution, affine transformation, marginalization
- convergence in distribution

It is a natural infinite-dimensional generalization of the class of Gaussian densities.

Quite flexible class for addressing key contemporary data challenges such as multivariate density estimation and regression.

Literature overview

[Walther, 2002] showed that for the problem of univariate log-concave density estimation, the MLE $\hat{f}_n$ exists and is computationally tractable.

When $d = 1$, there are faster algorithms for computing $\hat{f}_n$, including an active set algorithm called logcondens [Dümbgen and Rufibach, 2011] and a more recent constrained Newton method called cnmlcd [Liu and Wang, 2018].

[Cule et al., 2010] established that if $n \geq d + 1$, then $\hat{f}_n$ exists and is unique with probability 1, and is supported on the convex hull of the data points $X_1, \ldots, X_n$ (which is a convex polytope).

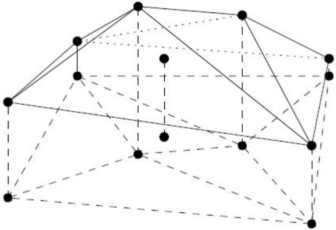

Figure: tent function from [Cule et al., 2010]

However, the computational time increases quickly with sample size and dimension.

Goal and setup

Our goal: use score matching to obtain computationally efficient yet flexible methods for log-concave density estimation.

Setup: Consider family of densities

$$f(x;\theta) \propto \exp\left(-s(x;\theta)\right),$$

where s is a univariate convex polynomial of degree $(2d+2)$

$$s(x;\theta) = \theta_{2d+1}x + \sum_{i=0}^{2d} \theta_i \cdot \frac{1}{(i+1)(i+2)} x^{i+2} \ .$$

Using score matching method [Hyvärinen, 2005] one can estimate the unknown parameter vector θ.

Define the score function

$$\psi(x;\theta) = \nabla_x \log f(x;\theta)$$

Define the objective function as

$$J(\theta) = \frac{1}{T}\sum_{t=1}^{T}[\psi'(x_t;\theta) + \frac{1}{2}\psi^2(x_t;\theta)]$$

$$\min_{\theta} J(\theta) \quad \text{s.t.}$$

$$\begin{cases} Q \succeq 0, \; Q = Q^\top, \\ \theta_i = \sum_{j+k=i} Q_{jk}, \quad i = 1, \dots 2d \\ \theta_0 = Q_{00} \end{cases}$$

Solve using CVXR

$$\left(H_i^{(d+1)}\right)_{jk} = \begin{cases} 1, \ j+k=i \\ 0, \ \text{otherwise.} \end{cases}$$

if $d = 2$ we have

$$H_0^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ H_1^3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ H_2^3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \ H_3^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\min_\theta J(\theta) \quad \text{s.t.}$$
$$\begin{cases} Q \succeq 0, \ Q = Q^\top, \\ \theta_i = \text{trace}(H_i^{(d+1)} \cdot Q) \end{cases}$$

Toy example

Let true density be

$$f(x;\theta) \propto \exp\left(-x^4 - 2x^2\right),$$

Perform $N = 100$ simulations for sample size 100

	Score-matching	Logcondens
MISE	0.03739245	0.05635883

Further steps

Consider family of densities

$$f(x;p) \propto \exp\left(-s(x)\right),$$

where s is any convex function.

Extension to regression models:

$$Y = \beta^{\top} X + \varepsilon,$$

where ε_i iid zero mean random errors with log-concave distribution.

Further steps

Consider family of densities

$$f(x;p) \propto \exp\left(-s(x)\right),$$

where s is any convex function.

Extension to regression models:

$$Y = \beta^\top X + \varepsilon,$$

where ε_i iid zero mean random errors with log-concave distribution.

References I

Cule, M., Samworth, R., and Stewart, M. (2010).
Maximum likelihood estimation of a multi-dimensional log-concave density.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(5):545–607.

Dümbgen, L. and Rufibach, K. (2011).
logcondens: Computations related to univariate log-concave density estimation.
Journal of Statistical Software, 39:1–28.

Grenander, U. (1956).
On the theory of mortality measurement: part ii.
Scandinavian Actuarial Journal, 1956(2):125–153.

References II

Hyvärinen, A. (2005).
Estimation of non-normalized statistical models by score matching.
Journal of Machine Learning Research, 6(4).

Liu, Y. and Wang, Y. (2018).
A fast algorithm for univariate log-concave density estimation.
Australian & New Zealand Journal of Statistics, 60(2):258–275.

Walther, G. (2002).
Detecting the presence of mixing with multiscale maximum likelihood.
Journal of the American Statistical Association, 97(458):508–513.