MAXIMUM LIKELIHOOD ESTIMATION OF A MULTI-DIMENSIONAL LOG-CONCAVE DENSITY THROUGH SOS-CONVEXITY

WEI HU, MAX CYTRYNBAUM, YJ CHOE

Abstract. We consider a tractable parametric relaxation of the log-concave maximum likelihood density estimation problem. Specifically, we let s(x) be a multivariate convex polynomial and consider densities of the form exp(-s(x)). While checking if a polynomial is convex is NP-hard in general, sos-convexity can be enforced using semi-definite programming. Log sos-concave maximum likelihood is formulated as a convex problem. We formulate an algorithm using projected stochastic gradient descent, in which biased gradient estimates are obtained through an MCMC sampling procedure that is efficient for log-concave densities. We motion towards the theoretical properties of our estimator, including consistency and asymptotic equivalence with the max-likelihood log-concave estimator discussed in Cule et al. (2008).

1. Introduction and Motivation

Consider the general shape constrained max-likelihood problem. We are given data $\{x_i\}_{i=1}^n$ and asked to find a density that maximizes the likelihood in some restricted class of densities $\mathcal{F}$. Formally, we seek

$$f^* = \operatorname*{argmax}_{f\in\mathcal{F}} \ell(f) = \operatorname*{argmax}_{f\in\mathcal{F}} \sum_{i=1}^{n} \log f(x_i) \qquad (1)$$

Log-concavity is a common shape constraint with a variety of well-studied properties. Densities in class have the form $f(x) = \exp(-s(x))$, where $s(x)$ is a convex function. Note that this nests several well known distributions, including normal, exponential, logistic, gamma (with shape parameter ≥ 1), and many more. Indeed, the standard multivariate normal distribution with parameters (μ, Σ), has a density proportional to $f(x) = \exp(-s(x))$, where $s(x) = \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$. Since $\Sigma^{-1} \succeq 0$, the quadratic form $x \to s(x)$ is clearly convex.

More importantly for our purposes, there exist efficient MCMC-based algorithms for sampling from log-concave densities. For these procedures, it can be shown that a Markov Chain with stationary distribution $f(x) = \exp(-s(x))$ is rapidly mixing. For more details and further references, see e.g. Narayanan and Rakhlin (2013) or Lovász and Vempala (2005).

Cule et al. (2008) proposes an algorithm for the general log-concave max-likelihood problem. They show that for $\mathcal{F} = \{\text{log-concave functions}\}$ the f^* in (1) above exists and is unique with probability 1. The paper also derives the form of the maximum likelihood estimator, showing that $x \to -s(x)$ is a piecewise-affine "tent function" supported on the convex hull of the data C_n. The density $f(x)$ is in general not smooth.

We consider a relaxation of this problem, restricting our attention to the class of log-concave functions where $x \to s(x)$ is an sos-convex polynomial. While checking a polynomial's convexity is in general strongly NP-hard (Ahmadi 2011), determining sos-convexity can be formulated as a

semi-definite program and can thus be checked in polynomial time. Using this insight, we propose an algorithm based on projected stochastic gradient descent, prove convergence, and demonstrate some of the properties of our estimator.

2. Convexity and SOS-Convexity

Note that this section closely follows the exposition in Amir Ali Ahmadi's 2011 MIT disseration (Ahmadi 2011).

As is well known, a function f defined on a convex set K is called *convex* if for all $x, y \in K$, $\lambda \in [0,1]$ we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Ahmadi (2011) shows that, even for polynomials, this condition can generally not be checked in polynomial time. Specifically, he gives the following result, answering a previously open problem posed by N.Z. Shor.

Theorem 2.1 (Ahmadi (2011) Theorem 2.1)**.** *Deciding convexity of degree-four polynomials is strongly NP-hard. This is true even when the polynomials are restricted to be homogeneous.*

This motivates the introduction of a relaxation of convexity, termed SOS or Sum-of-Squares convexity.

Given a polynomial matrix $U(x)$ over $\mathbb{R}^p$, we say that $U(x)$ is an SOS-matrix if there exists a factorization $U(x) = W(x)^T W(x)$, where W is also a polynomial matrix. As is well known, for a smooth function f, a necessary and sufficient condition for convexity is that the Hessian H_f is positive semidefinite $H_f \succeq 0$. Then we define

Definition 2.2 (SOS-Convexity)**.** *Let $f(x)$ be a polynomial over $\mathbb{R}^p$. We say that f is sos-convex if its Hessian is an sos-matrix i.e. there exists a polynomial matrix $W(x)$ such that*

$$H_f(x) = W(x)^T W(x)$$

Note that for such polynomials, we have for $y \in \mathbb{R}^p$

$$p(x,y) = y^T H_f y = y^T W(x)^T W(x) y = \|W(x)y\|_2^2$$

Then $p(x,y)$ is a sum-of-squares polynomial. In particular, we have $p \geq 0$ for all x, y, so $H_f \succeq 0$. Then clearly sos-convexity is a sufficient condition for convexity. In fact, it can be shown that we have the equivalent formulation of sos-convexity

Definition 2.3 (SOS-Convexity)**.** *Let $f(x)$ be a polynomial over $\mathbb{R}^p$. We say that f is sos-convex if for all $x, y \in \mathbb{R}^p$, the polynomial defined by*

$$p(x,y) = y^T H_f(x) y$$

is SOS.

Do there exist polynomials that are convex but not sos-convex? This question was answered in the affirmative by Ahmadi (2011). In fact we have

Theorem 2.4 (Ahmadi (2011))**.** *Consider polynomials of degree d over $\mathbb{R}^p$. Except in the cases $n = 1$, $d = 2$, and $(n, d) = (2, 4)$, there exists polynomials f such that f is convex but not sos-convex. In particular, for $p = 1$ (convex $\iff$ sos-convex).*

Nevertheless, the tractability of checking sos-convexity through semidefinite programming makes it an attractive relaxation of polynomial convexity. Using these definitions and intuition, we formulate the log sos-concave maximum likelihood problem.

3. The Log SOS-Concave Maximum Likelihood Estimation Problem

First, we introduce some notation. For $x \in \mathbb{R}^p$, we let $\mathbf{v}_d(x)$ denote the vector of all monomials in x i.e. $\mathbf{v}_d(x) = (x_1, \ldots, x_p, x_1^2, x_1x_2, \ldots, x_1^n, \ldots, x_p^n)$. We again consider the family of log sos-concave densities supported on a convex set $K \subseteq \mathbb{R}^p$:

$$p(\mathbf{x}) \propto \exp\left(-s(\mathbf{x})\right)$$

or

$$p(\mathbf{x}) = \frac{\exp(-s(\mathbf{x}))}{\int_K \exp(-s(\mathbf{t}))d\mathbf{t}}.$$

where $s(\mathbf{x})$ is an sos-convex polynomial. If we restrict the degree of s to be at most $2d$, then we can express s as

$$s(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in A_{2d}} \theta_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}.$$

where the $\boldsymbol{\alpha}$ are multi-indices on $\mathbb{N}^p$.

Given n i.i.d. samples $\mathbf{x}_1, \cdots, \mathbf{x}_n$ from a distribution $p(\mathbf{x}; \boldsymbol{\theta})$, the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{x}_i; \boldsymbol{\theta}) = \frac{\exp\left(-\boldsymbol{\theta}^T \sum_{i=1}^n \mathbf{v}_{2d}(\mathbf{x}_i)\right)}{(\int_K \exp(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}))d\mathbf{x})^n},$$

and then

$$-\frac{1}{n} \log L(\boldsymbol{\theta}) = \frac{1}{n} \boldsymbol{\theta}^T \sum_{i=1}^n \mathbf{v}_{2d}(\mathbf{x}_i) + \log \int_K \exp(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}))d\mathbf{x}.$$

So the maximum likelihood estimation of s (or equivalently, $\boldsymbol{\theta}$) can be summarized by the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \frac{1}{n} \boldsymbol{\theta}^T \sum_{i=1}^n \mathbf{v}_{2d}(\mathbf{x}_i) + \log \int_K \exp(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}))d\mathbf{x} \\ \text{s.t.} \quad & \boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}) \text{ is sos-convex.} \end{aligned} \tag{3.1}$$

Moreover we can show the following

Lemma 3.1 (Normalization)**.** *Suppose that θ^* is a solution to the above optimization problem. Then we have*

$$\int_K \exp\left(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x})\right) d\mathbf{x} = 1$$

i.e. the $\boldsymbol{\theta}$ returned by our optimization problem above automatically gives a normalized density.

Proof. Omitted.

Denote the above objective function by $g(\boldsymbol{\theta})$, which is a convex function. The gradient and Hessian of g are:

$$\begin{aligned}
\nabla g(\boldsymbol{\theta}) =& \frac{1}{n}\sum_{i=1}^n \mathbf{v}_{2d}(\mathbf{x}_i) + \frac{\int_K \exp(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}))(-\mathbf{v}_{2d}(\mathbf{x}))d\mathbf{x}}{\int_K \exp(-\boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}))d\mathbf{x}} \\
=& \frac{1}{n}\sum_{i=1}^n \mathbf{v}_{2d}(\mathbf{x}_i) - \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{v}_{2d}(\mathbf{X})), \qquad (3.2) \\
\nabla^2 g(\boldsymbol{\theta}) =& \mathbb{V}_{\boldsymbol{\theta}}(\mathbf{v}_{2d}(\mathbf{X})), \qquad (3.3)
\end{aligned}$$

where $\mathbf{X}$ is a random variable with distribution $p(\mathbf{x};\boldsymbol{\theta})$.

Note that since

$$\nabla^2 g(\boldsymbol{\theta}) = \mathbb{V}_{\boldsymbol{\theta}}(\mathbf{v}_{2d}(\mathbf{X})) \succeq 0 \qquad (3.4)$$

we immediately see that the problem defined above is convex.

3.1. **Projected Stochastic Gradient Descent.** We will solve (3.1) using projected stochastic gradient descent. This generates a sequence $\{\boldsymbol{\theta}_k\}_{k\geq 1}$ through the recursion:

$$\boldsymbol{\theta}_{k+1} \leftarrow P_{sos}(\boldsymbol{\theta}_k - \alpha_k(\nabla g(\boldsymbol{\theta}_k) + \xi_k)), k = 1, 2, \cdots \qquad (3.5)$$

where the initial point $\boldsymbol{\theta}_1$ is feasible for (3.1), $\{\alpha_k\}$ is a sequence of positive stepsizes, ξ_k is the (stochastic) error in the gradient evaluation, and $P_{sos}(\boldsymbol{\gamma})$ is the projection of $\boldsymbol{\gamma}$ onto the sos-convex cone $K_{sos}^{p,d} \subset \mathbb{R}^{\binom{p+d}{d}}$.

$$\begin{aligned}
P_{sos}(\boldsymbol{\gamma}): \qquad & \underset{\boldsymbol{\theta}}{\text{minimize}} \quad \|\boldsymbol{\theta} - \boldsymbol{\gamma}\|^2 \\
& \text{s.t.} \qquad \boldsymbol{\theta}^T \mathbf{v}_{2d}(\mathbf{x}) \text{ is sos-convex.}
\end{aligned} \qquad (3.6)$$

Our intuition for (3.5) is as follows: we would like to minimize the expression above, so we use gradient descent on $\boldsymbol{\theta}$. However, our descent steps $\boldsymbol{\theta}_k \to \boldsymbol{\theta}_{k+1}$ may take us outside of the sos-convex cone. Therefore, after each step, we project $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_k(\nabla g(\boldsymbol{\theta}_k) + \xi_k)$ back onto the sos-convex cone and continue.

3.2. **Projection onto $K_{sos}^{p,d}$ is an SDP.** Our problem (3.6) above is equivalent to

$$\begin{aligned} P_{sos}(\boldsymbol{\gamma}): \qquad & \underset{\boldsymbol{\theta},Q}{\text{minimize}} \quad \|\boldsymbol{\theta}-\boldsymbol{\gamma}\|^2 \\ & \text{s.t.} \qquad p(\boldsymbol{\theta},x)=\boldsymbol{\theta}^T\mathbf{v}_{2d}(x) \quad \text{is sos} \end{aligned} \tag{3.7}$$

We can show the following simple

Proposition 3.2. *The projection $P_{sos}(\boldsymbol{\gamma})$ onto the sos-convex cone $K_{sos}^{p,d}$ in* (3.7) *above can be formulated as an SDP.*

Proof. Consider the constraint that $p(\boldsymbol{\theta},x)$ is sos-convex. Let H denote the Hessian of this polynomial. Note that $\boldsymbol{\theta}\to H(\boldsymbol{\theta},x)$ is linear in $\boldsymbol{\theta}$. Then the constraint above is equivalent to

$$p(x,y,\theta)=y^TH(\boldsymbol{\theta},x)y \quad \text{is an sos-polynomial} \tag{3.8}$$

It is easy to show that this is equivalent to

$$\begin{aligned} & p(x,y,\theta)=y^TH(\boldsymbol{\theta},x)y=(\mathbf{v}_{2d}(x),y)^TQ(\mathbf{v}_{2d}(x),y) \\ & Q\succeq 0 \end{aligned} \tag{3.9}$$

i.e. $Q\in\mathbb{R}^{\binom{p+d}{d}+p}$ is some positive semidefinite matrix. Then the problem in (3.6) above can be reformulated as

$$\begin{aligned} P_{sos}(\boldsymbol{\gamma}): \qquad & \underset{\boldsymbol{\theta},Q}{\text{minimize}} \quad t \\ & \|\boldsymbol{\theta}-\boldsymbol{\gamma}\|^2<t \\ & y^TH(\boldsymbol{\theta},x)y=(\mathbf{v}_{2d}(x),y)^TQ(\mathbf{v}_{2d}(x),y) \\ & Q\succeq 0 \end{aligned} \tag{3.10}$$

It is routine to show that a quadratic constraint such as $\|\boldsymbol{\theta}-\boldsymbol{\gamma}\|^2<t$ above can be written as a linear matrix inequality (see e.g. Boyd (1998)). Therefore we get the final form

$$\begin{aligned} P_{sos}(\boldsymbol{\gamma}): \qquad & \underset{\boldsymbol{\theta},Q}{\text{minimize}} \quad t \\ & y^TH(\boldsymbol{\theta},x)y=(\mathbf{v}_{2d}(x),y)^TQ(\mathbf{v}_{2d}(x),y) \\ & \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}\succeq 0 \end{aligned} \tag{3.11}$$

Where the LMI constraint $P\succeq 0$ is equivalent to the quadratic constraint in (3.10) above. Since $\boldsymbol{\theta}\to y^TH(\boldsymbol{\theta},x)y$ is linear in $\boldsymbol{\theta}$ (and $Q\to a^TQa$ linear in Q) we have a linear objective function with LMI constraints, so this is an SDP. □

4. Algorithm Properties and Convergence

******
******

Wei - Insert updated algorithm properties document

******
******

5. Statistical Properties of the Max-Likelihood Log SOS-Concave Estimator

We discuss some of the statistical properties of our estimator. Here, we let K be a compact, convex set. We will assume that sos-convex polynomials are dense in continuous convex functions in the sense of $\|\cdot\|_{K,\infty}$. This assumption will be proved for the case $p = 1$.

Let f_n be the max-likelihood log-concave density returned by the algorithm in Cule et al. (2008). We show that

Theorem 5.1. *We have the following results*

(i) Convergence in likelihood values

$$g_d(\boldsymbol{\theta}^*) \longrightarrow \sum_{i=1}^{n} \log f_n(x_i) \qquad (d \to \infty) \tag{5.1}$$

That is, the objective value $g_d(\boldsymbol{\theta}^*)$ returned by our program (for polynomials up to degree d) approaches the objective value returned by the max-likelihood log-concave estimator as $d \to \infty$.

(ii) For any n, there exist a sequence of polynomials $p_n^m(\mathbf{x};\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in K_{sos}^p$ (we do not restrict degree) such that

$$p_n^m \xrightarrow{\mathcal{D}} f_n \qquad (m \to \infty) \tag{5.2}$$

Note that we will give the proof for general p, completing the denseness result for sos-convex polynomials in continuous convex functions only for the case $p = 1$.

Proof. Working on a compact, convex set K as above, we let SOSX denote the sos-convex polynomials, SMSX denote smooth strictly convex, SMX smooth convex, CTSX continuous convex. We will show the chain of denseness relations (in the sense of $\|\cdot\|_{\infty,K}$)

$$SOSX \overset{dense}{\subset} SMSX \overset{dense}{\subset} SMX \overset{dense}{\subset} CTSX \tag{5.3}$$

5.1. **Denseness of SOSX in SMSX.** Note that we are only able to give the proof for the case $p = 1$. Let $f \in SMSX$. In this case, we just have $K = [a, b]$ for some $a < b$. Then $f'' > 0$ on K. Using Stone-Weierstrass, let q_m be a sequence of polynomials $q_m \to f''$ in $\|\cdot\|_\infty$. Then there exists an M such that $m \geq M$ implies $q_m > 0$ on K. Restrict to this subset.

Define polynomials

$$p_m(x) = \int_a^x \left(\int_a^t q_m(s)ds \right) dt \tag{5.4}$$

Then clearly we have $\frac{\partial^2}{\partial x^2} p_m(x) = q_m(x)$. In particular, p_m is strictly convex on K, so p_m is an sos-convex polynomial by the results in section 2. Since $b - a < \infty$, it is easy to see that $p_m \to f$ uniformly on $[a, b]$ (one way to do this is to apply bounded convergence twice to the expression in

(5.4) above).

Therefore, we have shown that $\overline{SOSX} \supset SMSX$.

5.2. **Denseness of SMSX in SMX.** We want to show that smooth strictly convex functions are dense in smooth convex functions i.e. $\overline{SMSX} \supset SMX$. This is easy. Fix $\varepsilon > 0$. Let $f \in SMX$. Then $f_\varepsilon(x) := f(x) + \varepsilon\|x\|_2^2$ is strictly convex and

$$|f_\varepsilon(x) - f(x)| \leq \varepsilon\|x\|_2^2 \leq \varepsilon \cdot Diam(K)^2$$

for$x \in K$. Then $\|f_\varepsilon - f\|_\infty \to 0$ as $\varepsilon \to 0$, so we have the density result.

5.3. **Denseness of SMX in CTSX.** Here, we assume that $f \in CTSX(K)$. Extend f continuously to Suppose that f is such a function and let $\delta = d(K, K_1^c)$.

Convolution with a positive function preserves convexity. This is easily seen from the definition. Let f convex on $\mathbb{R}^p$ and $\varphi \geq 0$, then

$$(\varphi * f)(x) = \int_{\mathbb{R}^p} \varphi(t) f(x-t) dt$$

Each $f_t(x) := f(x-t)$ is convex, therefore the convolution above is convex.

Let φ be a smooth, compactly supported function (e.g. a bump function). For $\varepsilon > 0$, define a class of functions $\varphi_\varepsilon = \frac{1}{\varepsilon}\varphi(\frac{x}{\varepsilon})$. Fix f a continuous convex function. Then f is bounded on K. Define a class of functions $f_\varepsilon := (\varphi_\varepsilon * f)$. The standard arguments show that (i) f_ε is C^∞ for each ε (ii) $f_\varepsilon \to f$ uniformly on any compact set $A \subset\subset K$ and (iii) f_ε is convex for all ε. This will exhibit f as a limit point of SMX, showing that $\overline{SMX} \supset CTSX$.

Note that (iii) was shown above and essentially follows from the definition of convexity. Since φ is compactly supported and f continuous, the family of convolutions defined above converge for all $x \in \mathbb{R}^p$. First we need the following

Lemma 5.2. *Let $f \in CTSX(K)$, where $K \subset \mathbb{R}^p$ is compact and convex. Then*

$$\max_{g \in \partial f(x),\, x \in K} \|g\|_2 < \infty \tag{5.5}$$

is bounded.

Proof. Note that convexity guarantees the existence of subgradients at each point. In the smooth case, this just says that the gradient is bounded on K, which is clearly true since it is compact. The proof of the general statement is omitted but is relatively easy to show from the definition of a subgradient.

Lemma 5.3. *Suppose that f and K are as in the previous lemma. Then there exists an extension of f to a continuous and convex function $\overline{f}$ defined on $\mathbb{R}^p$.*

Proof. Omitted.

For (i), note that

$$(\varphi * f)(x) = \int_{\mathbb{R}^p} \varphi(t) f(x-t) dt = \int_{\mathbb{R}^p} \varphi(x-t) f(t) dt \tag{5.6}$$

so that whenever we need to take a derivative $\frac{\partial}{\partial x_\alpha}(\varphi * f)(x)$, we just pass the differentiation operator through the integral (formal justification using dominated convergence can be given since φ is compactly supported and f continuous). Thus, we just differentiate the integrand φ, which is C^∞ by assumption.

For (ii), let $\overline{f}$ be an extension of f from K to $\mathbb{R}^p$ as in (5.3). Choose $\varepsilon > 0$. Let $K_1 \supset\supset K$. We will show that $\overline{f}_\varepsilon \to \overline{f}$ on K. Note that we have $0 < \gamma = d(K, K_1^c)$. In what follows, we choose $\delta < \gamma$. Using uniform continuity of $\overline{f}$ on K_1, choose δ as above such that $x, y \in K_1$, $\|x-y\| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$. Finally, choose η such that $\int_{B(0,\delta)} \varphi_\eta(t) dt \geq 1 - \varepsilon$

Calculate as follows for $x \in K$

$$\begin{aligned} \overline{f}(x) - \overline{f}_\eta(x) &= \int_{\mathbb{R}^p} \overline{f}(x)\varphi_\eta(t) dt - \int_{\mathbb{R}^p} \overline{f}(x-t)\varphi_\eta(t) dt \\ &= \int_{\mathbb{R}^p} (\overline{f}(x) - \overline{f}(x-t))\varphi_\eta(t) dt = \int_{B(0,\delta)} (\overline{f}(x) - \overline{f}(x-t))\varphi_\eta(t) dt \\ &+ \int_{B(0,\delta)^c} (\overline{f}(x) - \overline{f}(x-t))\varphi_\eta(t) dt \end{aligned} \tag{5.7}$$

Consider the first term. By our choice of δ, in absolute value this is

$$\leq \int_{B(0,\delta)} |\overline{f}(x) - \overline{f}(x-t)| \varphi_\eta(t) dt \leq \varepsilon \int_{B(0,\delta)} \varphi_\eta(t) dt \leq \varepsilon \tag{5.8}$$

Next we consider the second integral at the end of (5.7) above. Define the set

$$A = \bigcup_{\eta < 1} \bigcup_{x \in K} (-\operatorname{Supp}(\varphi_\eta) + x) \tag{5.9}$$

and note that this set is bounded by compactness of K and (decreasing, nested) finite support of φ_η. A simple change of variables shows that the second integral can be rewritten as

$$\int_{B(0,\delta)^c} (\overline{f}(x) - \overline{f}(x-t))\varphi_\eta(t) dt = \int\limits_{-B(0,\delta)^c \cap \operatorname{Supp}(\varphi_\eta) + x} (\overline{f}(x) - \overline{f}(\xi))\varphi_\eta(x-\xi) d\xi \tag{5.10}$$

Since for any $x \in K$ and $\eta < 1$, we have that

$$-B(0,\delta)^c \cap \operatorname{Supp}(\varphi_\eta) + x \subset A \subset \overline{B}(0,R) \tag{5.11}$$

for some large R, then boundedness of f on $\overline{B}(0,R)$ (say by N), gives that the norm of the second integral is

$$\leq \int\limits_{-B(0,\delta)^c \cap \operatorname{Supp}(\varphi_\eta) + x} 2N \cdot \varphi_\eta(x-\xi) d\xi \leq \int_{B(0,\delta)^c} \varphi_\eta(t) dt \leq 2N\varepsilon \tag{5.12}$$

Then since $x \in K$ was arbitrary, we have shown that $\overline{f}_\varepsilon \to \overline{f}$ uniformly on K. In particular, $f \in CTSX(K)$ is the uniform limit of a sequence of smooth convex functions, so we have shown that $\overline{SMSX} \supset CTSX$. This finishes our density results, and we have shown that

$$SOSX \overset{dense}{\subset} SMSX \overset{dense}{\subset} SMX \overset{dense}{\subset} CTSX \tag{5.13}$$

in the uniform norm on a compact, convex set K. In particular, for $p = 1$, we have given a full proof that sos-convex polynomials are dense in continuous convex functions on a compact, convex set K.

5.4. **Distribution Convergence Argument.** Consider the shape-constrained maximum-likelihood estimator f_n from the Cule paper. In this section, we show that there exists a a sequence of log sos-concave densities p_n^m such that

$$p_n^m \xrightarrow{\mathcal{D}} f_n \qquad (m \to \infty) \tag{5.14}$$

We also show that, as $d \to \infty$, the objective value, i.e. the achieved likelihood, of our program converges to the max-likelihood among log-concave estimators (the objective value of Cule's program).

Our approach will be as follows. Given the convex hull of the data C_n, we will construct a set $K = C_n^\varepsilon$ such that $\mathcal{L}(C_n^\varepsilon \setminus C_n) = \varepsilon$, where $\mathcal{L}$ denotes Lebesgue measure. We will then extend the concave function s_n defining the Cule estimator ($f_n = \exp(s_n)$) to K in such a way that (i) our extension is convex and (ii) our extension is increasingly negative on ∂K.

Using our work above, let p_m be a sequence of sos concave functions that converge to our extension in $\|\cdot\|_\infty$. We will then argue that for any bounded function G on $\mathbb{R}^p$, we have

$$\int_{\mathbb{R}^p} G(x)\exp(p_m(x)) = \int_{K^c} G(x)\exp(p_m(x)) + \int_{K\setminus C_n} G(x)\exp(p_m(x)) + \int_{C_n} G(x)\exp(p_m(x))$$

We will show that (i) the first integral is small because $p_m(x) \approx -\infty$ on K^c. (ii) The second integral is small because $\mathcal{L}(C_n^\varepsilon \setminus C_n) \to 0$ as $\varepsilon \to 0$. Finally, (iii) the last integral converges to $\int_{C_n} G(x) f_n(x)$ because $\|p_m - s_n\|_{\infty, C_n} \to 0$. A rigorous formulation of this argument will show that for any bounded function G we have

$$\int_{\mathbb{R}^p} G(x)\exp(p_m(x))dx \to \int_{\mathbb{R}^p} G(x) f_n(x)dx \qquad (m \to \infty)$$

i.e. p_m converges to f_n in distribution.

Definition of K - Consider the tent function s_n from Cule's paper. $s_n = -\infty$ outside of C_n. Note that as a convex hull, we have $C_n = \{Ax \le b\}$ for some matrix $A \in \mathbb{R}^{n\times p}$ and $b \in \mathbb{R}^n$. Consider the set $C_n^\varepsilon = \{Ax \le b + \varepsilon 1\}$. Note that since C_n is of finite Lebesgue measure, then so is C_n^1 because this only depends on the a_i. Then we can apply elementary measure theory to show that

$$\mathcal{L}(C_n) = \mathcal{L}\left(\bigcap_{\varepsilon>0} C_n^\varepsilon\right) = \lim_{\varepsilon\to 0} \mathcal{L}(C_n^\varepsilon)$$

Therefore, for instance, we can make $\mathcal{L}(C_n^\varepsilon \setminus C_n)$ arbitrarily small by letting $\varepsilon \to 0$. Let $K_1 \supset C_n$ a polytope of the above form such that $\mathcal{L}(K_1 \setminus C_n) < \varepsilon$. From Cule, there exists a triangulation of C_n such that s_n is affine on each simplex in the triangulation.

Extend this triangulation to a triangulation of the polytope K_1 and extend s_n to $\overline{s_n}$ on K_1 by defining piecwise affine functions such that $\overline{s_n} = 0$ on ∂K_1. Since $Vol(K_1 \setminus C_n)$ is arbitrarily small and $\overline{s_n}$ is bounded on K_1, this extension does not affect the value of the integral

$$\int_{\mathbb{R}^p} G(x) \exp(s_n)$$

asymptotically as $\varepsilon = Vol(K_1 \setminus C_n) \to 0$. Therefore, wlog, we will assume from now on that s_n vanishes on ∂C_n.

Extension of s_n to K - Define $\widehat{s_n}$ such that $\widehat{s_n}(x) = s_n(x)$ for $x \in C_n$ and $\widehat{s_n}(x) = 0$ on C_n^c. Then s_n is continuous but is in general no longer concave. Define a function on K by

$$g(x) = \widehat{s_n}(x) - M \cdot d(x, C_n)$$

Claim - There exists an N such that $M \geq N$ implies that g is concave on K. Note that the set distance term is identically 0 on C_n and $\widehat{s_n}$ is identically 0 on C_n^c. Moreover, the set distance term is concave, since $x \to d(x, S)$ is a convex function whenever S is a convex set.

By concavity of the original Cule function s_n on C_n, for each point $x \in C_n$, the subgradient set $\partial s_n(x)$ is non-empty. In fact, from Cule's work we know that s_n has the form

$$\begin{aligned} s_n(x) &= \sum_k (a_k^T x + b_k) I(C_n^k) \quad (x \in C_n) \\ &= -\infty \quad else \end{aligned}$$

Where C_n^k is a triangulation of the convex hull C_n^k. We can show that

Lemma - For a function of the form above, $x \in C_n^k \Rightarrow a_k \in \partial s_n(x)$. The only non-trivial part of the argument is dealing with points $x \in \partial C_n^k$. Proof omitted.

Subgradient argument - To show that our extension $g(x)$ is concave, it suffices to show that $\partial g(x) \neq \emptyset$ for each $x \in K$. Then by concavity of the disjointly supported pieces of $g(x)$, it suffices to show that the subgradient condition is satisfied for each $x \in C_n$, $y \in K \setminus C_n$ and conversely.

Consider $x \in C_n$ and $y \in K \setminus C_n$. Let p_{y^*} denote the projection of y onto C_n. In our construction of $g(x)$, choose

$$M \geq N = \max_k \|a_k\|_2$$

i.e. over all vectors defining the piecewise affine function s_n. Then we can calculate as follows

$$\begin{aligned} g(y) - g(x) &= g(y) - g(p_{y^*}) + g(p_{y^*}) - g(x) = -M\|y - p_{y^*}\| + g(p_{y^*}) - g(x) \\ &\leq -M\|y - p_{y^*}\| + \partial g(x)^T(p_{y^*} - x) = -M\|y - p_{y^*}\| + \partial g(x)^T(p_{y^*} - y + y - x) \\ &\leq 0 + \partial g(x)^T(y - x) \end{aligned}$$

Note that the 1^{st} inequality follows by concavity of g on C_n, while the 2^{nd} inequality follows from Cauchy-Schwarz and our choice of M. A similar argument can be used to show that subgradients of g exist for $x \in K \setminus C_n$. Then g is concave on K.

By construction, for $M \geq N$ we have that g is a concave function. Note that $x \in \partial K \Rightarrow d(x, K) \geq \varepsilon$. Choose $K = C_n^{\varepsilon}$ as previously defined, where $M_\varepsilon \cdot \varepsilon = N_\varepsilon \to \infty$ as $\varepsilon \to 0$.

Convergence argument - We argue that $\int_{K^c} \exp(p_m) \to 0$ as $m \to \infty$. Let $x \in K^c$. Let π_x denote the projection of x onto C_n. Then there exists a point $x_0 \in \partial K \cap [\pi_x, x]$. Consider a function $\gamma(t) = p_m(\nu(t))$, where ν is a parametrization of the line $[\pi_x, x]$. Let $\nu(0) = \pi_x$ and $\nu(a) = x_0$. Then there exists a c such that

$$\begin{aligned} &- M\varepsilon = \gamma(a) - \gamma(0) = \gamma'(c)\|\pi_x - x_0\| \\ &\Rightarrow \gamma'(c) \leq \frac{-M\varepsilon}{\|\pi_x - x_0\|} \leq \frac{-M\varepsilon}{Diam(K)} \end{aligned}$$

Since γ is concave and $c \in [0, a]$, we then also have $\gamma'(a) \leq \frac{-M\varepsilon}{Diam(K)}$. Then Taylor implies that

$$\begin{aligned} p_m(x) = \gamma(1) &= \gamma(a) + \gamma'(a)\|x_0 - x\| + \frac{1}{2}\gamma''(d)\|x_0 - x\|^2 \\ &\leq -M\varepsilon + \frac{-M\varepsilon}{Diam(K)}\|x - x_0\| \leq \frac{-M\varepsilon}{Diam(K)}\|x - x^*\| \end{aligned}$$

Where x^* is the projection of x onto K. Let $\varphi_m = p_m + M\varepsilon$. Then we have

$$\int_{K^c} \exp(p_m) \leq \exp(-M\varepsilon) \int_{K^c} \exp(\varphi_m) \leq \exp(-M\varepsilon) \int_{K^c} \exp\left(\frac{-M\varepsilon}{Diam(K)}\|x - x_0\|\right)$$

We chose m such that $M(\varepsilon)\varepsilon \to \infty$ as $\varepsilon \to 0$. Then in particular for large enough m the integral is

$$\leq \int_{K^c} \exp\left(-\|x - x^*\|\right) = \int_{K^c} \exp\left(-d(x, K)\right)$$

This integral clearly converges since K is compact. Since $-M\varepsilon \to \infty$ as $\varepsilon \to 0$, for any bounded function $\|G\|_{\infty, \mathbb{R}^p} \leq B$ we have that

$$\int_{K^c} G \cdot \exp(p_m) \to 0 \quad (m \to \infty)$$

Summary of Integral Resuts - Note that by construction, we have $\mathcal{L}(C_n^{\varepsilon} \setminus C_n) \to 0$ as $\varepsilon \to 0$. Moreover, we used density of sos-convex polynomials to choose p_m such that $|p_m(x)| \leq |p_m(x) - g_m(x)| + |g_m(x)| \leq \varepsilon + 0 \leq 1$ on $C_n^{\varepsilon} \setminus C_n$. Therefore we easily have that

$$\int_{C_n^\varepsilon \setminus C_n} G \cdot \exp(p_m) \leq \int_{C_n^\varepsilon \setminus C_n} Be = Be \cdot \mathcal{L}(C_n^\varepsilon \setminus C_n) \to 0$$

For the final integral in the decomposition, we have

$$\int_{C_n} G \cdot \exp(p_m) \to \int_{C_n} G \cdot s_n$$

by bounded convergence on a finite measure space, with bound

$$\|G \exp(p_m)\|_{\infty, C_n} \leq B \cdot (\|s_n\|_{\infty, C_n} + 1)$$

Note that the latter bound is fixed for all m. Therefore, for any bounded function G we have that

$$\int_{\mathbb{R}^p} G \cdot \exp(p_m) \to \int_{\mathbb{R}^p} G \cdot \exp(s_n)$$

In particular

$$\int_{\mathbb{R}^p} \exp(p_m) \to \int_{\mathbb{R}^p} \exp(s_n) = 1$$

Let $z_m = \frac{\exp(p_m)}{\int_{\mathbb{R}^p} \exp(p_m)}$. Then the argument above shows that for any bounded function G

$$\int_{\mathbb{R}^p} G \cdot z_m \to \int_{\mathbb{R}^p} G \cdot f_n$$

So that we have convergence in distribution $z_m \to f_n$.

Convergence of objective function value - Consider a function z_m as constructed above such that $\|z_m - f_n\|_\infty \leq \varepsilon$. Then for data $\{x_i\}_{i=1}^n$, we must have in particular that $z_m(x_i) \geq f_n(x_i) - \varepsilon$. Therefore, if we let L denote the non-parametric likelihood, we have

$$L(z_m) \geq L(f_n) - n\varepsilon_m \to L(f_n) \quad (m \to \infty)$$

Since $z_m \in SOSX$, the function returned by our convex problem must do better, so in particular our objective value converges to Cule's objective value as we let $d \to \infty$ (our construction uses z_m of potentially arbitrarily high log-degrees).

Convergence of Our Estimator in Distribution - It remains to show that our estimator converges in distribution to the Cule estimator. Let our estimator for degree d be denoted by β_d. Then we know that $L(\beta_d) \to L(f_n)$ as $d \to \infty$. We somehow need to use convergence of likelihood values + the shape constraint that β_m is in particular log-concave.

6. Conclusion

Conclude here.

References

[1] Christakis, N., Fowler, J., Imbens, G., and Kalyanaraman, K., "An Empirical Model for Strategic Network Formation," *forthcoming*

[2] Mele, A., "A Structural Model of Segregation in Social Networks," *forthcoming*

[3] Hitsch, G., Hortacsu, A., and Ariely, D. "Matching and Sorting in Online Dating," *American Economic Review, 2010*

[4] Graham, B., Imbens, G., and Ridder, G. "Measuring the Effects of Segregation in the Presence of Social Spillovers: a Nonparametric Approach," *forthcoming*

[5] Fowler, F. and Christakis, N. (2007) "Cooperative Behavior Cascades in Human Social Networks," *Proceedings of the National Academy of Science, 2007*

[6] Jackson, M., "Social and Economic Networks," *Princeton University Press, 2008*