STATISTICAL PROPERTIES OF LOG SOS-CONCAVE DENSITY ESTIMATOR

MAX CYTRYNBAUM, WEI HU

1. Introduction

In this note, we will describe an assortment of results that we think could eventually lead to a nice analysis of the statistical properties of our estimator. The structure of this note is as follows:

(i) We show that sos-convex polynomials are dense in continuous convex functions in the sense of $\|\cdot\|_{\infty,K}$, where K is a fixed compact set.

(ii) Let f_n denote the Cule estimator for n data points (i.e. the max-likelihood log-concave estimator). We show that there exists a sequence of log sos-concave densities p_n^m on $\mathbb{R}^p$ (of potentially increasing associated degree) that converge to f_n in distribution on $\mathbb{R}^p$.

(iii) We show that this implies convergence of our objective function value to the likelihood value obtained by Cule's estimator f_n. We think that these facts will help us show that the estimator *returned by our algorithm* converges to Cule's estimator in distribution. If this is true, then the density returned by our algorithm converges in distribution to the min KL-divergence estimator as we let $n, d \to \infty$ i.e. it is asymptotically optimal.

2. Density of SOS-Convex Functions in Continuous Convex Functions

Fix a compact set $K \subset \mathbb{R}^p$ star-shaped with respect to $x_0 \in K$. All density results in this section are with respect to the norm $\|\cdot\|_{\infty,K}$. We begin by showing that sos-convex polynomials (SOSX) are dense in smooth, strictly convex functions on K (SMSX).

Fix $f \in SMSX$. Let $H_f(x)$ denote this function's Hessian. For $x \in K$ we have $H_f(x) \succ 0$. Therefore, we can take a Cholesky factorization

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

Note that since the component functions of H_f are smooth, recursively all of the component functions of L_f are in particular continuous (computing $L_{ij}(x)$ just involves adding, multiplying, and potentially taking the square root of other continuous functions).

By Stone-Weierstrass, $\mathbb{R}[x]_p$ is dense in $C(K)$, therefore we can find polynomials $L_{ij}^m(x) \in \mathbb{R}[x]_p$ with the property $\|L_{ij}^m(x) - L_{ij}(x)\|_\infty \to 0$ as $m \to \infty$ (in particular we can get simultaneous uniform convergence of the $L_{ij}^m \to L_{ij}$). Define a matrix

$$H_f^m(x) = L_f^m(x)L_f^m(x)^T$$

Then by the above work we have that

$$\|H_f^m - H_f\|_{\infty,K} \to 0 \qquad (m \to \infty)$$

Note that for each m, $H_f^m(x)$ is a polynomial matrix over $\mathbb{R}^p$ that factors as $H_f^m(x) = L_f^m(x)L_f^m(x)^T$, where the Cholesky factors *are also polynomial matrices*. Suppose that we can construct a polynomial p_m that has H_f^m as its Hessian i.e. $H_{p_m} = H_f^m$. Then by construction p_m is sos-convex.

By the above argument, we can find a sequence of polynomial matrices $H_f^m(x)$ such that

$$\|H_f^m(x) - H_f(x)\|_\infty \to 0 \quad (m \to \infty)$$

Build a sequence of polynomials $p_m(x)$ (of arbitrary degree) as follows. For $x_0 \in K$ (with respect to which K is star-shaped), define polynomials $p_m(x)$ such that

$$\begin{aligned}
&(1)\ p_m(x_0) = f(x_0)\\
&(2)\ \nabla p_m(x_0) = \nabla f(x_0)\\
&(3)\ H_{p_m}(x) = H_f^m(x) \quad \forall x \in K
\end{aligned}$$

For completeness, one way to do this is as follows. Suppose we want a function h such that $\nabla h = g$ and $h(x_0) = c$. Then we can set $h(x) = c + \int_0^1 g(\gamma_x(t)) \cdot \gamma_x'(t)\, dt$ for $x \in K$, where γ_x is the line segment $[x_0, x]$. This type of construction allows us to build a polynomial of the form required above.

Given a polynomial p_m that satisfies $(1) - (3)$ as above, it is simple to bound $\|f - p_m\|_\infty$. We include this calculation for the sake of completeness:

Convergence of functions from Hessian $\|\cdot\|_\infty$ convergence - Choose m such that we have $\|H_f^m - H_f\|_{\infty,K} < \epsilon$. Fix a path $\gamma(t)$ from $x_0 \to x$, which we will take to be the line segment $[x_0, x]$. Define a function $\phi(t) = (f - p_m)(\gamma(t))$. Then by Taylor's Theorem, we have

$$\begin{aligned}
&\phi(1) = \phi(0) + \phi'(0) + \frac{1}{2}\phi''(c) \quad c \in (0,1) \iff\\
&(f - p_m)(x) = (f - p_m)(x_0) + \nabla(f - p_m)(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T H_{f-p_m}(z)(x - x_0)
\end{aligned}$$

Where $z \in [x_0, x]$. By construction of our polynomials, the first two terms are identically 0. Moreover, the component functions $\gamma(x)$ of the Hessian $H_{f-p_m}(x)$ have $\|\gamma\|_\infty < \epsilon$. Then by Gershgorin's Theorem, we have an eigenvalue bound

$$\lambda(x) \leq \epsilon + \sum_{i \neq j} |\gamma_{ij}(x)| \leq p \cdot \epsilon$$

Since the Hessian is in particular Hermitian at each point, this gives a 2-norm bound on the matrix H_{f-p_m}. Using Cauchy-Schwarz, we have that

$$2|(f - p_m)(x)| \leq \|x - x_0\|_2 \|H_{f-p_m}(x)(x - x_0)\|_2 \leq \|x - x_0\|_2^2 \cdot p\epsilon \leq Diam(K)^2 \cdot p\epsilon$$

Then we have show that $\|H_f^m - H_f\|_\infty \to 0$ implies that $\|p_m - f(x)\|_\infty \to 0$. The p_m are all sos-convex, so we have shown that $\overline{SOS} \supset SMSX$.

Approximation of Smooth Convex Functions - 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 $\epsilon > 0$. Let $f \in SMX$. Then $f_\epsilon(x) := f(x) + \epsilon\|x\|_2^2$ is strictly convex and

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

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

Approximation of Continuous Convex Functions - Note that convolution with a positive function preserves convexity. This is easily seen from the definition. Let f convex and $\phi \geq 0$, then

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

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

Let ϕ be (for instance) the pdf of a Gaussian $\mathcal{N}(0, I)$ (not actually necessary, we just need the function we are convolving against to be smooth and integrate to 1). For $\epsilon > 0$, define a class of functions $\phi_\epsilon = \frac{1}{\epsilon}\phi(\frac{x}{\epsilon})$. Fix f a continuous convex function. If f is not bounded, redefine $f = 0$ on K^c. Define a class of functions $f_\epsilon := (\phi_\epsilon * f)$.

The standard arguments show that (i) f_ϵ is C^∞ for each ϵ (ii) $f_\epsilon \to f$ uniformly on any compact set and (iii) f_ϵ is convex for all ϵ. Then we have shown that f is a limit point of SMX, therefore $\overline{SMX} \supset CTSX$.

Sketch of convolution argument - I sketch the standard arguments for the sake of completeness. (iii) was shown above and basically follows from the definition of convexity.

For (i), note that

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

so that whenever we need to take a derivative $\frac{\partial}{\partial x_\alpha}(\phi * f)(x)$, we just pass the differentiation operator through the integral (formal justification for this can be provided) and differentiate ϕ, which is C^∞.

For (ii), rewrite the integral above as

$$\begin{aligned} f_\epsilon = (\phi_\epsilon * f)(x) &= \int_{\mathbb{R}^p} \phi_\epsilon(t) f(x-t) dt \\ &= \int_{B_\epsilon} \phi_\epsilon(t) f(x-t) dt + \int_{B_\epsilon^c} \phi_\epsilon(t) f(x-t) dt \end{aligned}$$

Where we choose B_ϵ a very small ball containing 0 of radius $h(\epsilon)$ such that (i) $\int_{B_\epsilon} \phi \approx 1$ and (ii) $\phi \approx 0$ on B_ϵ^c as well as (iii) $f(x-t) \approx f(x)$ for $t \in B_\epsilon$ (this is where we use continuity on the class $CTSX$). Then roughly speaking the above decomposition is

$$\begin{aligned} &\int_{B_\epsilon} \phi_\epsilon(t) f(x-t) dt + \int_{B_\epsilon^c} \phi_\epsilon(t) f(x-t) dt \\ &\approx f(x) \int_{B_\epsilon} \phi_\epsilon(t) dt + \int_{B_\epsilon^c} 0 \cdot f(x-t) dt \\ &\approx f(x) \cdot 1 + 0 = f(x) \end{aligned}$$

I claim that this gives uniform convergence $\|(\phi_\epsilon * f) - f\|_{\infty, K} \to 0$. Note that continuity of f on K implies uniform continuity, so the the approximation of the first integral in the above decomposition can be made uniform. f is uniformly bounded on $\mathbb{R}^p$, so the 2^{nd} integral above goes to 0 uniformly. This sketch can be made rigorous but the arguments are relatively standard.

Then we have shown that

$$SOSX \longrightarrow SMSX \longrightarrow SMX \longrightarrow CTSX$$

where the arrow denotes density (in the sense of $\|\cdot\|_{K,\infty}$ for a fixed K compact and star-shaped). In particular, we have shown that sos-convex polynomials are dense in continuous convex functions on a compact, convex set K.

3. Convergence in Distribution

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 that converge in distribution to the Cule estimator f_n.

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 $K = C_n^\epsilon$ such that $\mathcal{L}(C_n^\epsilon \setminus C_n) = \epsilon$, 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 such 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 (1) the first integral is small because loosely $p_m(x) \approx -\infty$ on K^c. (2) The second integral is small because $\mathcal{L}(C_n^\epsilon \setminus C_n) \to 0$ as $\epsilon \to 0$. Finally, (3) 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

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

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 $C_n^\epsilon = \{Ax \le b + \epsilon 1\}$. Note that C_n^1 is of finite Lebesgue measure since C_n is [HOW?]. Then we can apply elementary measure theory to show that

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

Therefore, for instance, we can make $\mathcal{L}(C_n^\epsilon \setminus C_n)$ arbitrarily small by letting $\epsilon \to 0$. Let $K_1 \supset C_n$ a polytope of the above form such that $\mathcal{L}(K_1 \setminus C_n) < \epsilon$. 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 $\epsilon = Vol(K_1 \setminus C_n) \to 0$. 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, $x \in C_n^k \implies a_k \in \partial s_n(x)$. 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 $x \in \partial K \implies d(x, K) \geq \epsilon$. Choose $K = C_n^\epsilon$ as previously defined, where $M_\epsilon \cdot \epsilon = N_\epsilon \to \infty$ as $\epsilon \to 0$.

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

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

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

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

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

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

We chose m such that $M(\epsilon)\epsilon \to \infty$ as $\epsilon \to 0$. 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 $-M\epsilon \to \infty$ as $\epsilon \to 0$, 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 $\mathcal{L}(C_n^\epsilon \setminus C_n) \to 0$ as $\epsilon \to 0$. Moreover, we used density of sos-convex polynomials to choose p_m such that $|p_m(x)| \leq |p_m(x) - g_m(x)| + |g_m(x)| \leq \epsilon + 0 \leq 1$ on $C_n^\epsilon \setminus C_n$. Therefore we easily have that

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

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 $\|z_m - f_n\|_\infty \leq \epsilon$. Then for data $\{x_i\}_{i=1}^n$, we must have in particular that $z_m(x_i) \geq f_n(x_i) - \epsilon$. Therefore, if we let L denote the non-parametric likelihood, we have

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

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.

Question - Does log-concavity + likelihood convergence imply convergence in distribution?

Things that we had trouble proving / omitted

** Show that we have a density relation $SOSX \longrightarrow SMSX$. This proof currently has problems.

** Prove or disprove that our estimator β_d converges in distribution to f_n.

** Need to show that $\mathcal{L}(C_n) < \infty \implies \mathcal{L}(C_n^1) < \infty$. Note that this immediately implies that $\epsilon \to \mathcal{L}(C_n^\epsilon \setminus C_n)$ is a continuous function by standard measure theory arguments - we just need the measure space to be finite.

**Argumentation for existence of appropriate subgradients on $K \setminus C_n$ omitted.